PL-TR-94-2211
Environmental Research Papers, No. 1156

DIAGNOSING CLOUDINESS FROM GLOBAL NUMERICAL WEATHER PREDICTION MODEL FORECASTS

Donald C. Norquist
H. Stuart Muench
Donald L. Aiken
Douglas C. Hahn

5 JULY 1994

PHILLIPS LABORATORY
Directorate of Geophysics
AIR FORCE MATERIEL COMMAND
HANSCOM AIR FORCE BASE, MA 01731-3010

This technical report has been reviewed and is approved for publication.

DONALD A. CHISHOLM , Chief
Satellite Analysis and Weather
Prrediction Branch

ROBERT A. McCLATCHEY
Director, Atmospheric Sciences Division

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the National Technical Information Service (NTIS).

If your address has changed, if you wish to be removed from the mailing list, of if the addressee is no longer employed by your organization, please notify PL/IM , 29 Randolph Road, Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	5 July 1994	Final Report Oct 91 to Sep 93

4. TITLE AND SUBTITLE
Diagnosing Cloudiness from Global Numerical Weather Prediction Model Forecasts

5. FUNDING NUMBERS
PE 62101F
PR 6670
TA GT

WU 32

6. AUTHOR(S)
Norquist, Donald C.
Muench, H. Stuart
Aiken, Donald L.
Hahn, Douglas C.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Phillips Laboratory (GPAB)
29 Randolph Road
Hanscom Air Force Base, Massachusetts 01731-3010

8. PERFORMING ORGANIZATION REPORT NUMBER
PL-TR-94-2211
ERP, No. 1156

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We investigated the utility of any information derivable from non-cloud numerical weather prediction (NWP) model forecasts in inferring layer cloud amount distributions. This effort involved identifying and preparing a suitable source of the predictand (cloud amount), generating and preparing a suitable source of the predictors (NWP variables and geographic information), and combining them to form diagnostic relationships in a model output statistics approach. Both AFGWC RTNEPH cloud analyses and Phillips Laboratory Global Spectral Model (PL GSM) NWP forecasts were rendered on a 125 km equal-area grid in three cloud deck regimes (high, middle, low). Two statistical methods [cloud curve algorithm (CCA), a univariate method, and multiple linear regression (MLR)] were used to relate the cloud amount to relative humidity (CCA) and to relative humidity and a large number of other NWP variables (MLR). We found that the CCA method preserves the sharpness of the cloud distribution while sacrificing skill, while MLR produced cloud diagnoses that were more skillful but less sharp. The methods fall short of the error level standards established by Air Force requirements, but show potential for useful cloud forecast skill upon further refinement.

14. SUBJECT TERMS
Numerical weather prediction
Cloud forecasting
Model output statistics
Multiple linear regression

15. NUMBER OF PAGES
152

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

Contents

Illustrations

Tables

Acknowledgement

We are grateful to Dr. Kenneth Mitchell and Mr. Kenneth Campana of the National Meteorological Center's Development Division for their recommendations on the use of the RTNEPH data. Dr. Thomas Nehrkorn of Atmospheric and Environmental Research, Inc. assisted us in putting together the PL-92 version of the PL GSM. Our thanks go to the European Centre for Medium-Range Weather Forecasting for supplying the global meteorological analysis data sets. Our branch chief, Donald Chisholm, gave us significant support and valuable input in the course of this project. Audrey Campana patiently performed the word processing tasks involved in the publication of this report.

Diagnosing Cloudiness from Global Numerical Weather Prediction Model Forecasts

1. INTRODUCTION

The U.S. Air Force Air Weather Service (AWS) is responsible for providing weather information to Air Force agencies on a global basis. A critical weather parameter required by many Air Force groups is the fractional cloud cover over an area of interest. This fractional cloud cover (which we will call "cloud amount" in this report) may be required over a particular region of interest in which a single value of cloud amount is needed for each of many subelements of this region. We refer to such an array of areal subelements of a larger area of interest as a grid, and the subelements as gridpoints (where each "point" represents the centroid of the spatial subelement).

The Air Force Global Weather Central (AFGWC) is the AWS agency responsible for producing and disseminating much of the weather information required in worldwide Air Force operations. AFGWC has specified requirements for global cloud

Received for publication 30 June 1994

analyses (cloud amounts at the present time) and global cloud forecasts (cloud amounts at future times) that are necessary to meet existing operational demands[1]. Global cloud analyses are required on an hourly basis on a 24 km resolution (that is, spacing between points) grid. Requirements for global cloud forecasts are divided into two categories: short-range (1-12 hour) forecasts are required every hour on a 24 km grid; long-range (13-120 hour) forecasts are required every 3 hours on a 95 km grid. Total cloud amount (that is, the fraction of the subelement covered by any clouds as viewed from above) is considered the highest priority, but the fractional coverage by each individual cloud layer is also important in the cloud analyses and forecasts.

The current cloud analysis procedure in use at AFGWC is called the real-time nephanalysis model, or RTNEPH. The RTNEPH has been described in detail by Kiess and Cox[2] and by Hamill et al[3]. The RTNEPH cloud analyses (which we will refer to simply as RTNEPH) are produced every three hours on a global grid that has a nominal grid resolution of 46 km. At each gridpoint, the RTNEPH contains information pertinent to cloudiness, including total cloud amount, and layer cloud amount (with altitudes of bases and tops) for up to four cloudy layers. The cloud amount specifications are derived from visible and infrared imagery sensors on polar orbiting Department of Defense (DOD) Defense Meteorological Satellite Program (DMSP) satellites, as well as conventional surface and upper air observations. Other information available at each grid location includes an estimate of the cloud type, what type of observation contributed to the cloud amount specification, and the age of the specification in hours (that is, the length of time since the last update of the amount).

The current techniques in use at AFGWC for short-range (0-9 hours) and long-range (0-48 hours) cloud forecasting are the High Resolution Cloud Prognosis (HRCP) model and the 5LAYER model respectively[4]. Both models use estimates of atmospheric moisture derived from RTNEPH cloud analyses as initial conditions. Using wind forecasts from the AFGWC Global Spectral Model[5], the cloud forecast models advect the moisture along a forecast trajectory, then convert the moisture forecasts to cloud forecasts. HRCP cloud forecasts are initiated every 3 hours, have

a time resolution of 3 hours, and a spatial resolution of nominally 46 km. HRCP is operated to produce forecasts in a limited area of the globe, usually coinciding with a recent swath of DMSP satellite imager data. 5LAYER produces cloud amount forecasts at three-hour intervals on a nominal 185 km grid out to 48 hours. In the tropics, AFGWC uses a cloud forecast model called TRONEW[4] which uses a simple diurnal persistence method (that is, cloud cover now will repeat itself in 24 hours).

It is clear that current capabilities at AFGWC do not meet the previously mentioned requirements. To meet these requirements, AFGWC has instituted a programmed upgrade to their cloud analysis and forecasting capabilities known as the Cloud Depiction and Forecasting System (CDFS) II[1]. It is expected that the new cloud analysis system developed under the Support of Environmental Requirements for Cloud Analysis and Archive (SERCAA) project[6] will replace RTNEPH and provide the required cloud analysis capability. Current CDFS II planning does not include a similar specific replacement for HRCP and 5LAYER.[1] However, it is expected that an improved cloud analysis will benefit forecasts from these two models. In addition, CDFS II may be able to take advantage of products of ongoing and future research in alternative cloud forecast methods. This research will study the inclusion of cloud formation and decay processes in trajectory models, satellite-derived trending methods in short-term forecasts, and cloud diagnosis from numerical weather prediction (NWP) model forecasts for longer-term cloud amount forecasts. This last approach is the topic of the present report.

Recent DOD decision-making has delegated all global NWP forecast support for DOD operations to the Navy's Fleet Numerical Oceanographic Center (FNOC). As such, FNOC will routinely provide global NWP forecast data to AFGWC to meet their various large-scale, weather-related requirements. Thus, the most straightforward way of conducting routine large-scale, long-term cloud forecasting support at AFGWC that depends on global NWP inputs is in a stand-alone "post-processor" mode. This approach requires no changes to the global data analysis and forecast system at FNOC. As is true of all operational centers' global NWP models, cloudiness is not carried as an explicitly predicted (prognostic) variable. A cloud diagnostic method,

rather than a prognostic approach which would require changes to FNOC's NWP model, deserves investigation as a first attempt to use global NWP forecasting to benefit global cloud forecasting.

Some prior research has been conducted in the use of NWP model forecasts of noncloud meteorological parameters to diagnose cloud amounts at forecast times. Geleyn[7] developed a quadratic function of forecast relative humidity (RH) in association with a tunable parameter known as the critical relative humidity (a function of the model's vertical coordinate only) in an attempt to diagnose cloud amount from RH at any model time step. The motivation for this, and most other attempts at cloud diagnosis in NWP models, was to specify cloudiness for the radiative parameterization. Slingo[8] made major changes to the Geleyn procedure, implemented in the European Centre for Medium Range Forecasting (ECMWF) NWP model, introducing a more complex scheme. This scheme combines separate algorithms for convective and nonconvective clouds, a constant critical relative humidity, and a complex handling of low clouds depending on vertical motion and static stability. A number of reports[9,10,11] have described adaptations of the Slingo[8] scheme for other NWP models.

Mitchell and Hahn[12] tested several simple empirical cloud-humidity relations and encountered large biases that varied both in space and forecast projection, no matter which scheme was used. A major problem uncovered was the tendency for the humidity distribution to evolve towards a "model-preferred" vertical and latitudinal distribution different than the initial or "normal" condition, due to inadequacies in the moist physics. To compensate for this problem, the authors devised a procedure to produce cloud-humidity relations by matching the cumulative frequency distribution for observed cloud cover to the cumulative frequency distribution for model gridpoint humidity, with separate relations for different levels and forecast times. Independently, Rikus and Hart[13] developed a similar procedure. The procedure is quite effective at removing the biases and producing stable distributions of cloud cover. However, the procedure performed only slightly better than the simple diagnostic schemes in root-mean-square (RMS) error (RMSE), and only in some

circumstances could compete with the 5LAYER model skill. The relatively modest skill was blamed on the poor initial humidity analysis used, the error introduced in the pre- and post-processing steps, and the geographic dependence of the cloud-humidity relationship that they were unable to account for in their study. They found that the global spectral model (GSM) cloud forecasts suffered an initial disadvantage since they are initialized by a humidity field derived wholly independently of the RTNEPH (which is the verification). They recommended that for short-term GSM cloud forecasts, the RTNEPH be used solely or in concert with conventional observations for humidity initialization. They also recommended the use of the trending approach (adding the inferred forecast change of cloud amount to the initial RTNEPH amount) when the initial GSM moisture state is based on anything but RTNEPH.

Kvamsto[14] used the University of Bergen version of the Norwegian Meteorological Institute mesoscale model (50 km, 10-layer, surface to 200 hPa) to make forecasts of 21 and 27 hours to compare model humidity fields to cloud fields derived from polar-orbiting satellites for a storm case. The diagrams presented show an alarming scatter of gridpoint humidity for a given cloud cover. Even though the model does include ice physics, the correlation appears better for clouds with tops below 3 km. The humidity values for a given cloud cover were noted to be "drier" at 1500 UTC (mid-afternoon) than 0900 UTC (mid-morning) suggesting a diurnal effect. The scatter in the points (± 10 to 20 percent) may have been due to phase error in the forecasts of small features or may indicate the atmosphere is capable of producing different cloud covers for a given gridpoint humidity. He concludes that the short-term humidity forecast is more realistic than the humidity analysis at the corresponding time. This implies that humidity analyses cannot be used for verification of model simulations of humidity. The humidity field becomes more realistic during the integration period as the model gets enough time to structure this humidity field.

Sheu and Curry[15] performed a point-by-point verification of three different empirical schemes: Slingo[8], Mitchell and Hahn[12], and Smith[16]. The authors chose an

area in the north Atlantic (10-50W, 40-60N) in January 1979 (FGGE Special Observing Period) and obtained special ECMWF analysis data for a 1.875-degree grid on constant pressure levels. Cloud information was obtained from the AFGWC 3DNEPH (forerunner of the RTNEPH) data base. The performance of the three-cloud specification schemes was rather disappointing. The correlation between observed and specified cloud cover was about 0.0 at 1000 hPa, rising to about + 0.4 at 700 and 500 hPa and dropping to near 0.0 at 200 hPa, with small differences between the schemes. One might conclude: (a) the cloud data contain substantial noise; (b) the analyzed humidity fields are inaccurate (at least on a point-wise basis); (c) gridpoint humidity is insufficient to accurately specify cloud cover. However, the techniques were accurate at specifying the monthly mean area-average cloud cover.

Trapnell[17] performed an extensive point verification of experimental cloud forecasts using the AFGWC GSM and the Mitchell and Hahn[12] scheme (referred to here as the "cloud curve algorithm," or "CCA" method) using AFGWC RTNEPH data as "truth." Tests were run during several periods of 1991 and 1992, with variations in stratification by region and vertical motion, a vertical humidity compaction, a diurnal adjustment and other attempts at improving performance. In general, forecast scores were better than in the development tests, largely due to the improvements in the humidity initialization procedure prior to forecast. However, 6-hour forecasts were only slightly better than at 12 or 18 hours, indicating performance similar to that found by Shue and Curry[15] for specifying cloud from a zero-hour forecast. Areal mean cloud amount biases were smaller than those produced by the "5LAYER" model. In some cases, skill score statistics were competitive with those of 5LAYER, but were never clearly superior. Trapnell[17] felt that a fairer comparison between CCA in the GSM and 5LAYER would require (1) executing the GSM at higher spatial resolution, (2) outputting the GSM RH forecast directly on GSM sigma-level surfaces, and (3) utilizing the RTNEPH directly in deriving the GSM initial moisture analysis.

Zivkovic and Louis[18] developed an empirical cloud specification technique that easily accommodates multiple variables. A "cluster analysis" is performed by

obtaining principal components of variables, such as vertical profiles of temperature, humidity, and wind. The components explaining the most variance are retained and become the basis for the clusters, which are interpreted as weather "types," each associated with a different distribution of cloud cover frequency. A limited test was performed using NMC analyzed data (100 km resolution) on four January days, with cloud cover derived from surface observations at synoptic times. The resulting specifications of cloud cover showed some improvement over a simple quadratic formulation using RH and critical RH.

Bao[19] compared forecasts of clouds from the PSU/NCAR mesoscale model at 80 km resolution 3DNEPH cloud amounts in low, middle, and high altitude category designations. Both a diagnostic (cloud fraction is a simple linear function of RH) and prognostic (predicted cloud water, then this used in a relationship to deduce fractional cloud) were tested. Forecasts were evaluated against 3DNEPH interpolated horizontally to the 80 km model grid. Random and maximum overlap specifications were used to get low, middle, and high clouds from 15 3DNEPH layers. His major findings: (a) too much overcast is predicted by the model with either scheme, (b) conventional overlap (maximum and random) assumptions on 3DNEPH layer amounts cannot recreate the "total cloud" given in 3DNEPH, so the model clouds should not be expected to do so, (c) the model underestimates the area of high fractional coverage (this would suggest that the model cloud diagnosis could benefit from a scheme that corrects for humidity biases, like the CCA scheme), (d) the prognostic method underestimates nonovercast cloud fractions more often than the diagnostic scheme does, (e) instantaneous release of latent heat in nonprecipitating gridpoints which just become overcast (without neighboring points doing so) can lead to small-scale noise in such prognostic models, and (f) though the cloud fraction to cloud water relation could be reversed to specify initial cloud water in such a prognostic model, grid boxes that are less than saturated will quickly lose their cloud water early in the forecast. Bao[19] makes several recommendations: (1) there should be a consistent definition of cloud amount for both the observations and the NWP model predictions, (2) a statistical relationship between cloud amount and model

variables needs to be carefully established by the perfect-prediction method or by the model output statistics method, (3) a treatment of the hydrological processes, which gives a consistently good prediction of both precipitation and associated cloud activity, is needed in the NWP model, and (4) reliable sources of [cloud] data for initialization and forecast verification are necessary.

Krishnamurti et al.[20] compared diagnostic cloud forecasts from three GSMs with International Satellite Cloud Climatology Project (ISCCP) cloud amount data sets. Each model used its own cloud diagnosis scheme, all of which were fixed temporally but included spatial variations (horizontal and vertical). They noticed a spin-up in cloud amount from a zero specification to a specification based on the initial humidity analysis. The models generated their own humidity fields and the diagnostic schemes were tuned to those spun-up humidity states. They attributed the growing difference between the model clouds and the ISCCP clouds to: (1) the errors in the forecast of the cloud-producing synoptic systems, and (2) the spin-up of the models as the forecast period continues.

Wu[21] developed an empirical relationship between model forecast RH and diagnosed cloud such that the latter would result in optimal agreement with observed outgoing longwave radiation (OLR). To do this, he used a nonlinear programming technique known as active set algorithm to minimize a functional proportional to the sum of the squares of the vertically-summed difference of the model-produced upward radiative flux and the observed top of the atmosphere upward flux (OLR). The model-produced flux is dependent on the amount of cloud that has a direct transmittance to space (is not overlapped by higher cloud) and its temperature. This cloud amount is in turn dependent on the overlap scheme used (here, random) and the diagnostic scheme used (here, a hyperbolic tangent function of RH only, with two parameters). The method then uses the RH from the model and observed OLR and derives the values of the parameters that ensures the best fit of observed and inferred OLR. This then sets the cloud diagnosis (RH to cloud) parameters empirically. The parameters were derived separately for each of four temperature categories rather than by low, middle, and high or model sigma level categories. This empirical relationship is

derived over a set of NWP model gridpoints (to which observed OLR has been analyzed) which, in this case, was a limited region encompassing just the U.S., Mexico, and surrounding waters. The derivation may take place any time there is sufficient coverage to be a data assimilation technique. It potentially could improve the model's cloud diagnosis at the beginning or at any time during an analysis cycle. This would presumably improve the model's ensuing short-term forecast, which means a better first-guess for the next analysis, and thus, a better analysis. He saw a reduction (over a Smagorinsky diagnosis) of OLR error to background levels in the winter, but only a slight reduction in summer. He attributed this to using RH only as the predictor, which would not account well for convection dominant in summer.

Cianciolo[22] applied the Model Output Statistics (MOS) approach to a one-year sample of AFGWC GSM forecasts and corresponding RTNEPH cloud analyses. The goal was to build MOS models (relationships between NWP forecast quantities and cloudiness) for limited regions (RTNEPH boxes) to produce 3-, 6-, and 9-hour high resolution (46 km) total cloud amount forecasts. Different MOS models were developed for different combinations of RTNEPH boxes, seasons, and times of day. The study found that the MOS-based forecasts outperformed persistence with respect to the overall reduction in mean square error, but did not display the sharpness (percent of points which are clear, nearly clear, nearly overcast or overcast) seen in the cloud analyses.

Many of the studies reviewed above used RH as the sole or primary indicator of cloud amount. In our study, a major goal was to investigate the utility of using any NWP model predicted or diagnosed quantity as a supplement to RH as a cloudiness predictor. In practice, we limited our choice of potential predictors from the NWP model to those predicted or diagnosed quantities that we felt might in any way be linked with cloud amount distribution. Our purpose was to see if a combination of predictors may be able to reduce the random error (such as root-mean-square error) associated with cloud diagnosis schemes which rely solely or primarily on RH. The ultimate motivation of this research was to determine if it is possible to attain appreciable cloud forecast skill through a diagnostic method that relies on a subset

of information available from an NWP model that runs independently of the cloud analysis procedure. Accordingly, the independent cloud analyses (here, RTNEPH) were used to develop and verify the diagnostic scheme, but were not used in any way to influence the NWP forecasts. It may be left for further study to determine if any schemes so developed may be further improved by allowing the cloud analysis to influence the forecasts (at any time during the forecast procedure) of the noncloud predictor variables.

2. METHODOLOGY

As stated in the introduction, the goal of this project was to determine functional relationships between cloud amount and NWP model forecast quantities that may be used to diagnose clouds from future NWP model forecasts. Because there are no known analytical functional relationships, we knew that it would be necessary to develop empirical relationships, since we have ruled out numerical prognostic methods for reasons stated in the introduction. This puts us into the realm of statistical weather forecasting, the other branch of objective weather forecasting. Glahn[23] has defined statistical weather forecasting as "... forecasting through the use of a formal statistical analysis of the data, with the results of that analysis being clearly stated." He described the three general methods of applications of statistical models as the classical method, the perfect prog method, and the model output statistics method. Each method consists of a development phase and an application phase. The development phase is the process of collection and analysis of the data representing the variable to be predicted (predictand) and the variable(s) used to make the prediction (predictor(s)) in order to form an empirical relationship of the predictand as a function of one or more predictors. The data set used to develop the relationship is called the dependent data set, and should be inclusive enough to involve a wide enough range of predictand-predictor value combinations to account for future occurrences. The application phase is the forecast process in which the relationship is used in conjunction with values of the chosen predictors to forecast

values of the predictand at future times. The sample of data over which the statistical relationship is applied for forecasting is called the independent data set. In the classical method, the dependent data set is made up of predictand values at forecast times and predictor values available at observation times. The relationship obtained is applied to independent predictor values to obtain corresponding predictand forecasts. In the perfect prog method, the dependent data set consists of predictand (a variable not available from NWP model forecasts) values at observation times and predictor values (variables predicted by an NWP model) available at observation times. The relationship obtained is applied to independent predictor values from an NWP model at the desired forecast times to obtain corresponding predictand forecasts valid at those forecast times. Finally, in the model output statistics (MOS) method, the dependent data is composed of predictand values observed at times that correspond to the forecast times of the predictor values, which are taken from NWP model forecasts valid at those forecast times. The resulting relationship is applied to future NWP model forecasts of those predictors to get predictand forecasts.

2.1 The Dynamic Model Output Statistics Method

We tried to allow for the planned operating environment at AFGWC under CDFS II in selecting a general method of statistical forecasting. Since reputable NWP model forecasts would be available, we excluded the classical method, in which the method itself is the only predictive mechanism in the process. As the name implies, the perfect prog method works best when the NWP model forecast predictor values most closely resemble the observed values. Unfortunately, all forecast models create both random errors (those varying from point to point and from case to case) and systematic errors (those having the same nature over a spatial or temporal domain). Using the perfect prog method, these improperly specified values of the predictors at forecast times would yield similarly erroneous values of the predictand through the relationship derived using observations of both predictor and predictand variables. This factor can be overcome to some degree using the MOS method, especially in

dealing with systematic error. If the MOS relationships are derived using the same NWP model that will be used to generate future forecasts of the predictor variables, then the MOS technique can allow for NWP model biases if the development of the relationships is limited to the spatial and temporal domain in which the biases occur. The MOS method will relate whatever values of the predictors that the NWP model gives it to observed values of the predictand, regardless of the NWP model bias in that particular region and season. Ironically, if the NWP model forecast model is improved to reduce seasonal or regional biases, the resulting MOS-based predictand forecasts would be expected to degrade because of the dependence of the MOS relationships on the nature of the NWP model's forecasts. In this case, the MOS relationships would have to be rederived for the improved version of the NWP model.

Typically, the MOS method has been developed for and applied to individual locations (such as forecast centers in pre-selected cities), using NWP model forecasts over domains that include those locations. This is a way of using the large-scale model forecasts, whose resolution cannot fully represent the localized conditions of a single point, to predict meteorological conditions for individual localities. In this case, it is necessary to collect a dependent data set of predictors and predictands separately for each locality for each season. In order to allow for inter-annual fluctuations in weather at a certain location, the dependent data set should consist of a robust temporal sample over at least several years.

Other applications of the MOS method have developed predictand-predictor relationships over limited regions over which climatic conditions are fairly homogeneous, especially where the target area for the forecasting procedure is regional (tens to hundreds of kilometers on a side) rather than a locality. In this case, a larger sample of gridpoints from within the region, with their varying local conditions, will constitute a greater sample of data at any one time and thus may decrease the need for a larger temporal sample in the dependent data set.

We have attempted to account for the likelihood of a changing NWP forecast model in our application of the MOS method to cloud forecasting. In our design, our MOS relationships are developed from a limited temporal sample of recent NWP

model forecasts and concurrent cloud analyses. We chose the 10-day sample of twice-daily NWP model forecasts and RTNEPH cloud analyses, immediately prior to the date of the initial time of the target forecasts, as the predictand-predictor relationship development period. Our spatial domain for this study was the entire Northern Hemisphere. We acknowledge that this domain is not climatologically homogeneous, and that NWP model variable-cloud amount relationships may be different in the different climatic regimes that exist within the hemisphere. However, establishing the entire hemisphere as the development domain allows us to baseline the capability of the prediction technique for later determination of the benefit of regional application of the technique. Furthermore, it ensures that the 10-day sample will include all weather types that may be encountered in the forecast period. A recent and changing development period makes it likely that the forecasting method will encounter the same types of weather-cloud combinations found in the development sample, possibly improving its chances for success. Finally, the fact that the development period is constantly changing with the target forecast period allows for predictand-predictor relationships to account automatically for changes in the NWP forecast model.

2.2 Multiple Linear Regression

In this project, we chose multiple linear regression (MLR) as the specific method in using NWP model forecasts to predict cloud amount. Several factors were considered in making this choice. First, the National Weather Service[24] (NWS) has used MLR quite successfully in developing statistical equations to predict surface weather variables (including total cloud cover) based on NWP forecast output. Next, this method can be useful when a number of weather factors can affect the predictand through physical relationships not well formulated, which appeared likely to be the situation in cloud prediction. Lastly, the method is relatively simple and direct, systematically reducing the variance, in this case, cloud error-squared, through algorithms commercially available in statistical computer software.

The typical MLR procedure begins by developing an n x n covariance matrix, consisting of n-1 potential predictors and the predictand. The matrix is formed by computing all possible cross-products from each data set (in our case, each gridpoint, each time), summing over all data sets, dividing by the number of sums, and subtracting the cross-product of the averages. The next step is to select a subset of potential predictors to specify the predictand and compute the appropriate variables for the regression equation. In the IMSL/STAT/LIBRARY software used in this study, several selection strategies are available, including simple stepwise (backwards or forwards) and a best combination routine. The best combination routine proved to be computationally impractical for the 99 predictors used in this study, and following the suggestion of Murphy[25], a forward stepwise selection was used. In this procedure, the predictor with the best correlation to the predictand is found. Then the correlation between the "best" predictor and other predictors is removed from the other predictors and the next best predictor is sought. This process is continued until a desired number of predictors for the subset is obtained. In meteorological applications, investigators find very small increase in explained variance (reduction error) when the number of predictors in the subset increases beyond about 10 (see Figure 1). In this study, we chose 20 as the desired limit, hoping to capture the maximum predictive information.

For analysis purposes, the important results of the MLR process are the variance, reduction in variance (correlation coefficient), and the predictors chosen. In addition, the MLR routine also provides mean values and the regression slopes for chosen predictors. These quantities are used in assembling the linear prediction equations to be used in the independent tests.

2.3 Selection of Predictands

In our study, the primary predictand was RTNEPH cloud amount. We recognize that studies[3,26] have shown that the RTNEPH cloud amount representation departs from natural cloud distributions in several ways. For example, the RTNEPH tends

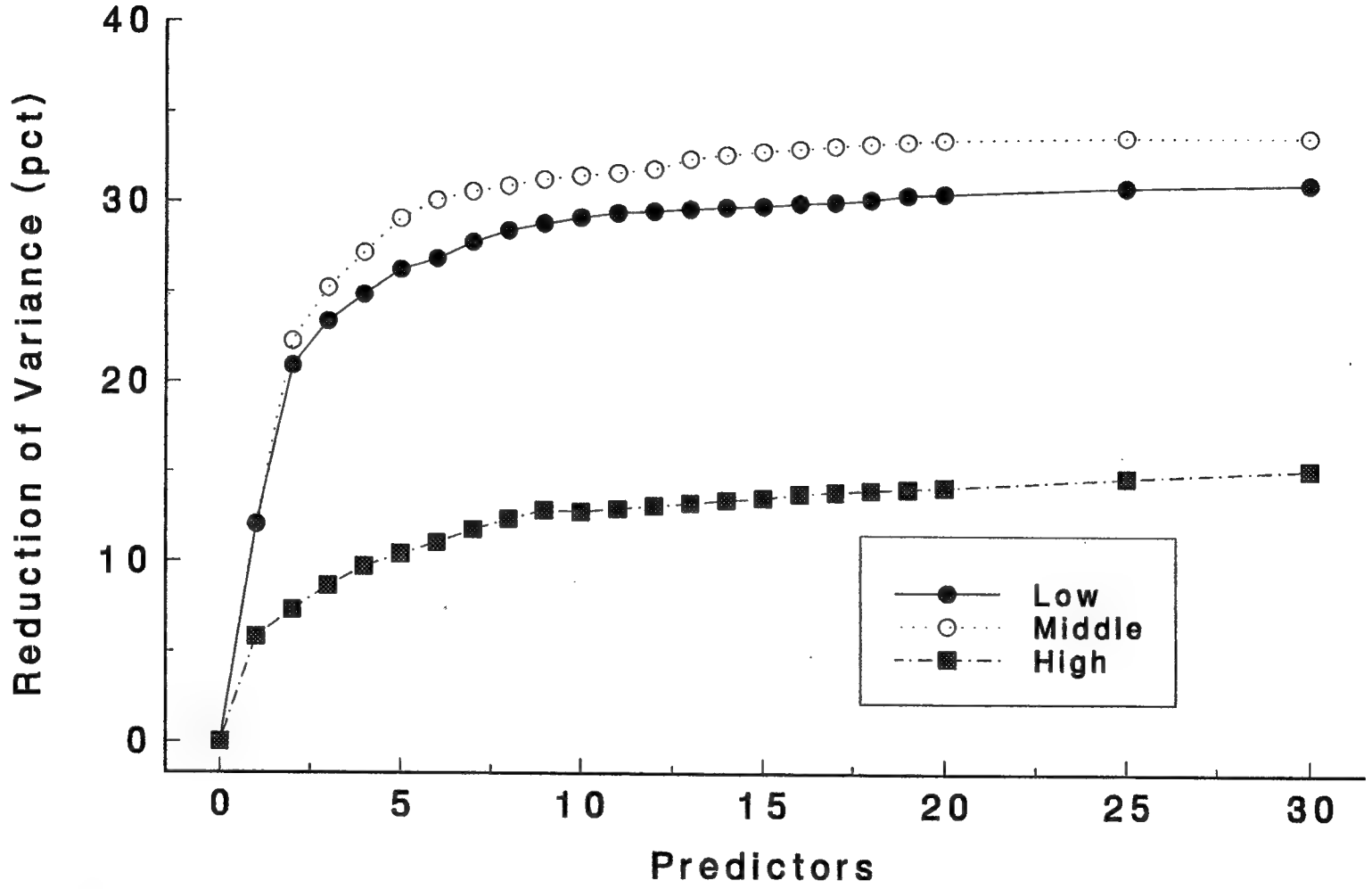

Figure 1. Percent Reduction of Variance in Multiple Linear Regression as a Function of the Number of Chosen Predictors. The Case Shown Represents the Results of a Single 10-Day Development Period in January 1991 for the Middle Cloud Deck Used in This Study.

to underestimate the amount of low clouds in the wintertime high latitudes, because of the difficulty in distinguishing cloudiness from the cold, snow or ice-covered ground, or ocean surface. Still, the RTNEPH cloud analysis accurately represents the important weather-related cloud features, and has a spatial resolution exceeding any current global NWP model resolution. This fact allows the RTNEPH to represent significant detail in inhomogeneous cloud scenes--for example, distinct frontal bands followed by well-defined dry slots and by regions of post-frontal instability marked by sectors of convective cloudiness. For large-scale applications, where forecasts for cloudiness of subelements of ~ 100 km on a side are desired, the resolution and accuracy of the RTNEPH is sufficient for development and verification of forecasting methods.

In addition to RTNEPH cloud amount, we used the difference in cloud amount between cloud curve algorithm estimates and RTNEPH as a secondary predictand in separately derived predictand-predictor relationships. The basic method for deriving cloud curve algorithm (CCA) cloud amounts has been described by Mitchell and Hahn[12] and Trapnell[17]. Briefly, the method is as follows. A sample of at least one week of global NWP model RH forecasts and RTNEPH cloud analyses corresponding to the valid time of the forecasts are used to form the dependent data set. After representing both the RH forecast and cloud amount analyses on the same three-dimensional grid, the frequency of occurrence of each 1 percent value (between 0 and 100 percent) is computed over the sample for both RH and cloud amount (usually separately for each vertical layer category; Trapnell[17] also developed the method separately for different geographic regions). The next step is to calculate the cumulative frequency of occurrence, which for each 1 percent category is the sum of the percent frequencies of occurrence for all 1 percent categories less than or equal to that category. Finally, for each 1 percent category, the cumulative frequency of occurrence of cloud amount is matched with the equivalent value of cumulative percent frequency of occurrence of RH. Thus, each 1 percent cloud amount category "finds" a 1 percent RH category that corresponds to it. This is the value of RH that is to indicate the corresponding cloud amount in the application. The highest value

of RH derived in this way corresponding to zero cloud amount is referred to as the critical relative humidity (RH_c). In application, any forecast value of RH less than or equal to RH_c is translated into zero cloud amount. Similarly, the value of RH that represents 100 percent cloud amount is called the maximum RH (RH_m), and in application any RH $\geq RH_m$ is translated into 100 percent cloud amount. For all RH values in between RH_c and RH_m, there is a one-to-one correspondence between RH and cloud amount. This relationship can be represented in the form of a table, or fit with an algebraic function relationship. The cloud amount vs. RH empirical relationships (referred to as CCA curves) so derived are applied to the ensuing NWP forecast to produce a cloud amount forecast, which we refer to as the CCA cloud amount. It is the difference between this and the verifying RTNEPH cloud amount that we used as a secondary predictor in separate MLR methodology development and application experiments.

There are several reasons for considering CCA-RTNEPH cloud amount differences as a predictand. First, the range of values of the differences are twice that of the range of RTNEPH cloud amounts. We felt that the MLR methodology may be able to reproduce both positive and negative values better than just being restricted to positive values. Secondly, the CCA scheme automatically accounts for NWP model drift in humidity at each atmospheric level and each forecast valid time, because separate cloud-humidity relationships are derived and applied for each. CCA cloud amount forecasts displayed a near-zero bias when the verification was computed over the same geographic area for which the relationships were derived and applied[17]. Finally, we felt that perhaps, since the average CCA-RTNEPH difference is about zero, no bias (when computed over the entire development domain) would be introduced in converting the predicted CCA-RTNEPH values into cloud amount.

While total area mean cloud amounts showed little bias as a result of using the CCA forecast method, the point-by-point accuracy of gridpoint cloud amount was disappointing in the earlier studies.[15,17] While in some cases, the point-by-point accuracy statistics (for example, RMS error) of the CCA forecasts were competitive with those of 5LAYER, they could in no cases (different cloud layers, different

seasons, different forecast lengths, different geographic areas) be judged clearly and consistently better than those of 5LAYER. We believe that this is due to the fact that, as the forecast continues, cloud amounts inferred from forecast relative humidity at individual points are not well correlated with RTNEPH cloud amounts. It was our intention in this project to identify forecast variables (predicted or diagnosed) or combinations of such variables that are more closely correlated with RTNEPH cloud amounts or CCA-RTNEPH cloud amount differences.

2.4 Selection of Predictors

In selecting *a priori* the predictor variables that we felt would be good indicators of cloudiness, we used four somewhat different approaches. The actual choices were based in part on the study by Cianciolo[22] and in part on experience in NWP and local forecasting. The multiple linear regression techniques seek independent specification information of predictors as they are added, so there is an advantage to seeking a diverse assortment of predictors not related to each other.

The first approach was to consider the humidity-cloud relation. The earliest approach[27] simply proposed a linear relation between the grid-volume-average RH and the horizontal cloud cover. Physically, there is no argument that shows that such a relation should be linear nor is there an obvious reason that RH should be better than other humidity parameters. Thus, we included in the predictor list, the volume-average RH to the 1st, 2nd and 4th powers as well as condensation-pressure-spread, the lifted-condensation-distance, and the precipitable water. A natural assumption would be that the cloud cover would correlate best with the humidity of the NWP model sigma layer with the highest humidity, so that humidity value was included as a predictor. On the other hand, because there may be a significant amount of noise or error in the humidities for individual sigma layers at individual gridpoints, the deck-average values were included. Relative humidity with respect to ice was also included, but since this model has no ice microphysics, it would not likely add useful information. Finally, we included the CCA forecast cloud amount diagnosed in the

manner discussed in Section 3.6 as a potential humidity-based predictor.

A major problem developing NWP model moist physics routines is that we do not have good measurements of volume-averaged humidity for verification. Typically, the moist physics routines are "tuned" to produce reasonable patterns of humidity and rainfall, verified against very localized radiosonde and rain gauge measurements. Undoubtedly, there are residual humidity errors, and a presumption was made that by using basic model dynamic variables known to be related to moisture, some compensation could be made for systematic error in the development and decay of humidity patterns. Predictors thus chosen included vorticity, divergence, temperature, the three components of motion, horizontal vorticity and temperature advection, and three-dimensional moisture convergence. From the moist physics routines, 6-hour rates of surface evaporation, stratiform precipitation and convecting precipitation were included as those parameters that directly impact model humidity. Except for those moist physics parameters, the model parameters are instantaneous (representative of a model time step, about five minutes). As such, they may be temporally unrepresentative. Some reduction of noise can be made by time averaging. The easiest way to do so was to include predictor sets at prediction time and 6 hours earlier (if the average does reduce variance, the regression will choose both times, giving similar coefficients). A more comprehensive approach would require extensive modifications to the NWP model data output procedures, counter to our goal of developing a procedure easily applied to operational models.

Simple examination of cloud imagery from satellites reveals great complexity in cloud patterns, much of which is at scale sizes too small to be resolved by the global NWP model. The small-scale clouds can only result from vertical motion and small-scale disturbances, which, for our purposes, we will call "turbulence." If turbulence activity is consistent in time and space, or is directly correlated to humidity, the effect will be properly included in the regression equations. Experience indicates that turbulence is highly variable in time and space and could create situations where different cloud cover scenarios could exist for a given layer-average humidity. Models are now being implemented that carry turbulent kinetic energy as a variable (e.g.,

NWS Eta model) but the present version of the GSM does not. As an alternative, some predictors were derived from model variables that were believed to be related to the amplitude of turbulent motions. Static stability values were computed as deck averages and for the σ layer having the maximum humidity in the deck. Mountain wave turbulence has been found to spill over to adjacent flatter areas, so we also included the minimum stability over three decks in a 3X3 gridpoint array centered on the predictand point. Wind shear represents a source for turbulent energy and was computed for each deck. The deck averaged wind speed was included as well as the speed in the lowest model layer. Further, maximum speed in a 3X3X3 gridpoint array was computed as well as a 3X3 measure of wind speed in the lowest layer of the model. The natural logarithm of the Richardson number for each deck was also computed and a 3X3X3 minimum computed. Latent heat release also represents a source of turbulent energy and thus moist static stability for each deck was computed. Finally, a 3X3 maximum of convective precipitation was used, which might indicate turbulent energy propagating out from convective areas.

In recognition of the fact that NWP model characteristics vary by region and time of day, and that RTNEPH cloudiness can be regionally and temporally dependent, we included a set of geographic predictors. An advantage of these predictors is that they are largely independent of the highly variant dynamic predictors and thus may add useful information even though the actual correlation may be small. The geographic predictors included latitude, sine and cosine of latitude, and sine and cosine of longitude. The percentage of each equal-area gridpoint that exists over water surfaces was included. Solar parameters and terrain variability may add information about turbulence conditions. For solar effects, we included solar zenith angle, cosine of zenith angle, hours since sunrise and hours since sunset. The equal area grid terrain height was added, as it might differ from the smoother model terrain, and a terrain variability was included based on standard deviation of 5 km average terrain height about the gridpoint mean. A wind-terrain variable was computed by multiplying the lowest model layer wind speed times the terrain variability.

Our list of predictors are shown in Table 1. All combinations of the above, including values of multi-layer predictors at the three cloud decks, total 99 predictors and one predictand available for ingest into the MLR scheme for each gridpoint in each deck.

2.5 Use of 12-Hour Forecasts to Study Forecast Method Potential

Two sources of random error are present in cloud amount diagnosis from noncloud NWP forecasts. The diagnosis procedure introduces random errors, even if the NWP forecasts are error free. This is because similar cloud amount vertical profiles can be associated with significantly different vertical profiles of observed noncloud meteorological parameters. Then, as the NWP model forecast proceeds forward in time, the random errors of the forecasted noncloud variables increase. To estimate the best-case capability of the cloud forecasting methods, we sought to isolate just the diagnosis procedure error.

In reality, even zero hour NWP model forecasts depart from observed reality, partially because of the difficulty in reconciling the spatial scale differences between model grids and observations. In fact, meteorological objective analysis schemes seek to arrive at the "optimal blend" between observed values and background field (a previous short-term forecast from the NWP model, valid at the time of the observations). Because of the local influences of the observations, the meteorological analysis has more spatial variance than the background field, and this greater degree of variance is unresolvable by the NWP model and is seen as small-scale "noise." Initialization procedures are executed on the analyses to remove the noise while preserving the information in the analysis in the scales that the model can resolve. However, many existing operational initialization schemes do not treat the analyzed absolute humidity distribution, so it returns much of the small-scale variance introduced in the analysis process. It isn't until the NWP model integration process is underway that this unresolved humidity variance is removed (often resulting in spurious precipitation in locations that the model sees as being too moist) and the

Table 1. List and Description of the "100-Predictors" Used in the Selection of Multi-Linear-Regression Cloud Predictors.

No.	Name	Description
1	VORD6	Vorticity, predictand deck average, forecast t-6
2	DIVD6	Divergence, predictand deck average, forecast t-6
3	TMPD6	Temperature, predictand deck average, forecast t-6
4	PRWD6	Precipitable water, predictand deck average, forecast t-6
5	RHUD6	Relative humidity, predictand deck average, forecast t-6
6	OMGD6	Vertical velocity, predictand deck average, forecast t-6
7	STBD6	d(theta)/d(z), predictand deck average, forecast t-6
8	SPDD6	Wind speed, predictand deck average, forecast t-6
9	SHRD6	Wind shear, predictand deck average, forecast t-6
10	VADD6	Vorticity advection, predictand deck average, forecast t-6
11	TADD6	Temperature advection, predictand deck average, forecast t-6
12	QADD6	3-D humidity div., predictand deck average, forecast t-6
13	CPSD6	Condens. pres. deficit, predictand deck average, forecast t-6
14	MSTD6	d(theta-e)/d(z), predictand deck average, forecast t-6
15	UCMD6	West wind component, predictand deck average, forecast t-6
16	VCMD6	South wind component, predictand deck average, forecast t-6
17	RHXC6	Maximum RH within predictand deck, forecast t-6
18	RHAC6	RH at layer above maximum RH (see #17), forecast t-6
19	TMPC6	Temperature at maximum RH (see #17), forecast t-6
20	STBC6	d(theta)/d(z) at maximun RH (see#17), forecast t-6
21	SFCP6	Surface pressure (not sea-level), forecast t-6
22	RFST6	6-hr stratiform surface precipitation, forecast t-6
23	RFCV6	6-hr convective surface precipitation, forecast t-6
24	EVAP6	6-hr surface evaporation, forecast t-6
25	SPDB6	Surface-layer wind speed, forecast t-6
26	VORH2	Vorticity, high deck average, forecast t-0
27	VORM2	Vorticity, middle deck average, forecast t-0
28	VORL2	Vorticity, low deck average, forecast t-0
29	DIVH2	Divergence, high deck average, forecast t-0
30	DIVM2	Divergence, middle deck average, forecast t-0
31	DIVL2	Divergence, low deck average, forecast t-0
32	RHUH2	Relative humidity (RH), high deck average, forecast t-0
33	RHUM2	Relative humidity, middle deck average, forecast t-0
34	RHUL2	Relative humidity, low deck average, forecast t-0
35	OMGH2	Vertical velocity, high deck average, forecast t-0
36	OMGM2	Vertical velocity, middle deck average, forecast t-0
37	OMGL2	Vertical velocity, low deck average, forecast t-0
38	STBH2	d(theta)/d(z), high deck average, forecast t-0
39	STBM2	d(theta)/d(z), middle deck average, forecast t-0
40	STBL2	d(theta)/d(z), low deck average, forecast t-0
41	SPDH2	Wind speed, high deck average, forecast t-0
42	SPDM2	Wind speed, middle deck average, forecast t-0
43	SPDL2	Wind speed, low deck average, forecast t-0
44	SHRH2	Wind shear, high deck average, 12- forecast t-0
45	SHRM2	Wind shear, middle deck average, 12-forecast t-0
46	SHRL2	Wind shear, low deck average, 12- forecast t-0
47	RHCH2	Maximum RH within high deck, 12- forecast t-0
48	RHCM2	Maximum RH within middle deck, 12 forecast t-0
49	RHCL2	Maximum RH within low deck, forecast t-0
50	TMPD2	Temperature, predictand deck average, forecast t-0

Table 1. (cont.) List and Description of the "100-Predictors" Used in the Selection of Multi-Linear-Regression Cloud Predictors.

No.	Name	Description
51	PRWD2	Precipitable water, predictand deck average, forecast t-0
52	VADD2	Vorticity advection, predictand deck average, forecast t-0
53	TADD2	Temperature advection, predictand deck average, forecast t-0
54	QADD2	3-D humidity div., predictand deck average, forecast t-0
55	CPSD2	Condens. pres. deficit, predictand deck average, forecast t-0
56	MSTD2	d(theta-e)/d(z), predictand deck average, forecast t-0
57	UCMD2	West wind component, predictand deck average, forecast t-0
58	VCMD2	South wind component, predictand deck average, forecast t-0
59	RHAC2	RH for level above RH-max, predictand deck, forecast t-0
60	TMPC2	Temperature at maximum RH (see #17), forecast t-0
61	STBC2	d(theta)/d(z) at maximun RH (see#17), forecast) t-0
62	SFCP2	Surface pressure (not sea-level), forecast t-0
63	RFST2	6-hr stratiform surface precipitation, forecast t-0
64	RFCV2	6-hr convective surface precipitation, forecast t-0
65	EVAP2	6-hr surface evaporation, forecast t-0
66	SPDB2	Surface-layer wind speed, forecast t-0
67	RH2C2	Maximum-RH-squared within predictand deck, forecast t-0
68	RH4C2	Maximum-RH-fourth within predictand deck, forecast t-0
69	CCAC2	CCA cloud forecast, predictand deck, forecast-t-0
70	RHIC2	RH wrt ice at RH maximum, predictand deck, forecast t-0
71	LCDC2	Lifted-cond.-dist. at RH maximum, pred. deck, forecast t-0
72	LRIC2	Ln(Ri.-Number) at RH maximum. predictand deck, forecast t-0
73	GSLAT	Latitude (Gaussian grid, GS)
74	SGSLA	Sine of Latitude
75	CGSLA	Cosine of Latitude
76	SGSLO	Sine of Longitude
77	CGSLO	Cosine of Longitude
78	ZENA2	Solar zenith angle, forecast t-0
79	CZEN2	Cosine of solar zenith angle, forecast t-0
80	HRSS2	Hours of sunshine before forecast t-0
81	HRDK2	Hours of darkness before forecast t-0
82	SFCHT	Surface terrain height (9-pt ave., 1/8 mesh data)
83	SDVHT	Standard deviation of surface terrain height
84	PCH2O	Percent of surface that is water (from 1/64 mesh data)
85	DZ/DX	Eastward gradient of terrain height
86	DZ/DY	Northward gradient of terrain height
87	LRN92	3x3x3 (ijk) minimum of Ln(Ri.-Number), forecast t-0
88	STN92	3x3x3 minimum of d(theta)/d(z), deck average, forecast t-0
89	SHX92	3x3x3 maximum of vertical shear, deck average, forecast t-0
90	SPX92	3x3x3 maximum of wind speed, deck average, forecast t-0
91	RCX92	3x3 maximum of 6-hr convective rainfall, forecast t-0
92	SBX92	3x3 maximum of surface layer wind speed, forecast t-0
93	SVX92	3x3 maximum of surface-speed-times-terrain-var., frcst t-0
94	WBLL2	Surface wind times terrain gradient, forecast t-0
95	ABTV2	Minimum of terrain ht. or wind/stability height, frcst t-0
96	RH2D2	RH-squared, predictand deck average, forecast t-0
97	RH4D2	RH-fourth, predictand deck average, forecast t-0
98	RH2D6	RH-squared, predictand deck average, forecast t-6
99	RH4D6	RH-fourth, predictand deck average, forecast t-6
100	CLDOB	Predictand, observed RTNEPH deck cloud cover, forecast t-0

spatial humidity scales match those of the mass and motion fields. For this reason, Kvamsto[14] and others state that using a short-term NWP forecast is a better choice for correlation with gridded representations of cloud cover than is the NWP analysis. Mitchell and Hahn[12] showed that the organization of the large-scale humidity field shows a better spatial correspondence with observed cloud patterns in the short-term forecast than in the initial analysis.

We chose a 12-hour forecast to examine the best-case capability of the cloud diagnosis method--that is, to best isolate the diagnosis error. We felt that this time might represent the best compromise between the model's organization of weather systems/patterns and forecast variable accuracy. Using the results from the 12-hour forecasts as a baseline, we could then measure the degradation in skill out to 48 hours of forecast time resulting from growing NWP model forecast error.

3. DATA SETS AND EXPERIMENTAL DESIGN

For this study, we selected the months of January and July 1991, and limited the domain of study to the Northern Hemisphere. We expected that this time-space domain would provide ample coverage of seasonal extremes (winter and summer) over a variety of geographic regimes. Seasonally, we expected that summertime and wintertime applications would indicate the range of the forecasting methods' capabilities. Geographically, we felt that applying the method over an entire hemisphere would be representative of the ultimate global application desired for this type of cloud forecasting method.

3.1 The Transform Grid

To reconcile the horizontal representation of RTNEPH cloud analyses with the NWP model forecasts, we chose the ECMWF equal-area grid[28] as the common grid, which we refer to as the transform grid. This grid has the same latitude locations as the Gaussian grid (the grid on which the spectral NWP model computes its nonlinear

terms in grid space), but the number of longitudes at each latitude are chosen so that each gridpoint at any latitude represents approximately the same surface area of the earth. Figure 2 shows a comparison of the Gaussian grid with the equal-area grid. We chose to use the equal-area grid as the transform grid because the NWP model, run at a triangular 106 wave (T106) resolution, is more coarsely resolved (each equal-area gridpoint is about 125 km on a side) than the RTNEPH analyses (gridpoints about 46 km on a side). For this reason, the NWP model grid scale is the limiting spatial scale. Thus, rather than interpolating NWP model forecasts to the RTNEPH grid (where the interpolated fields would not contain the spatial scales associated with the more highly resolved observed cloudiness), we averaged the RTNEPH to the transform grid. In our case, a 3X3 array of RTNEPH gridpoints extends, on average, just 6.5 km beyond the boundaries of a single transform gridpoint. To match each 3X3 RTNEPH array with the corresponding transform gridpoint, we found the RTNEPH gridpoint whose centroid lies closest to the transform gridpoint centroid, then used this RTNEPH gridpoint and its contiguous neighbors as the 3X3 array.

Ideally, it would be desirable to be able to use RTNEPH cloud amounts to specify not only the horizontal cloud fraction in each transform grid subelement, but the vertical fraction of the subelement occupied by clouds as well. If the base and top altitudes of each reported cloud layer at a given RTNEPH gridpoint were known to be highly accurate, this would be possible. This is because at each NWP model grid location, the NWP model layer interface altitudes are diagnosed from the temperature and surface pressure forecasts. Theoretically, one could use this information to identify which NWP model layer(s) were occupied by RTNEPH cloud, and the vertical fraction to which they are cloud-filled. In practice, however, the degree of accuracy to which the RTNEPH cloud base and top altitudes are known[3] cannot support the identification of even the corresponding NWP model layer. The finest vertical discrimination that we felt was justifiable for RTNEPH cloud reports is the specification of high, middle, and low cloud decks. Using the NWP model $\sigma = p/p_{sfc}$ vertical coordinates, we defined the bases of each of the three designated decks as $\sigma = 0.45$ for high, $\sigma = 0.80$ for middle, and $\sigma = 0.99$ for low. For both months, we used

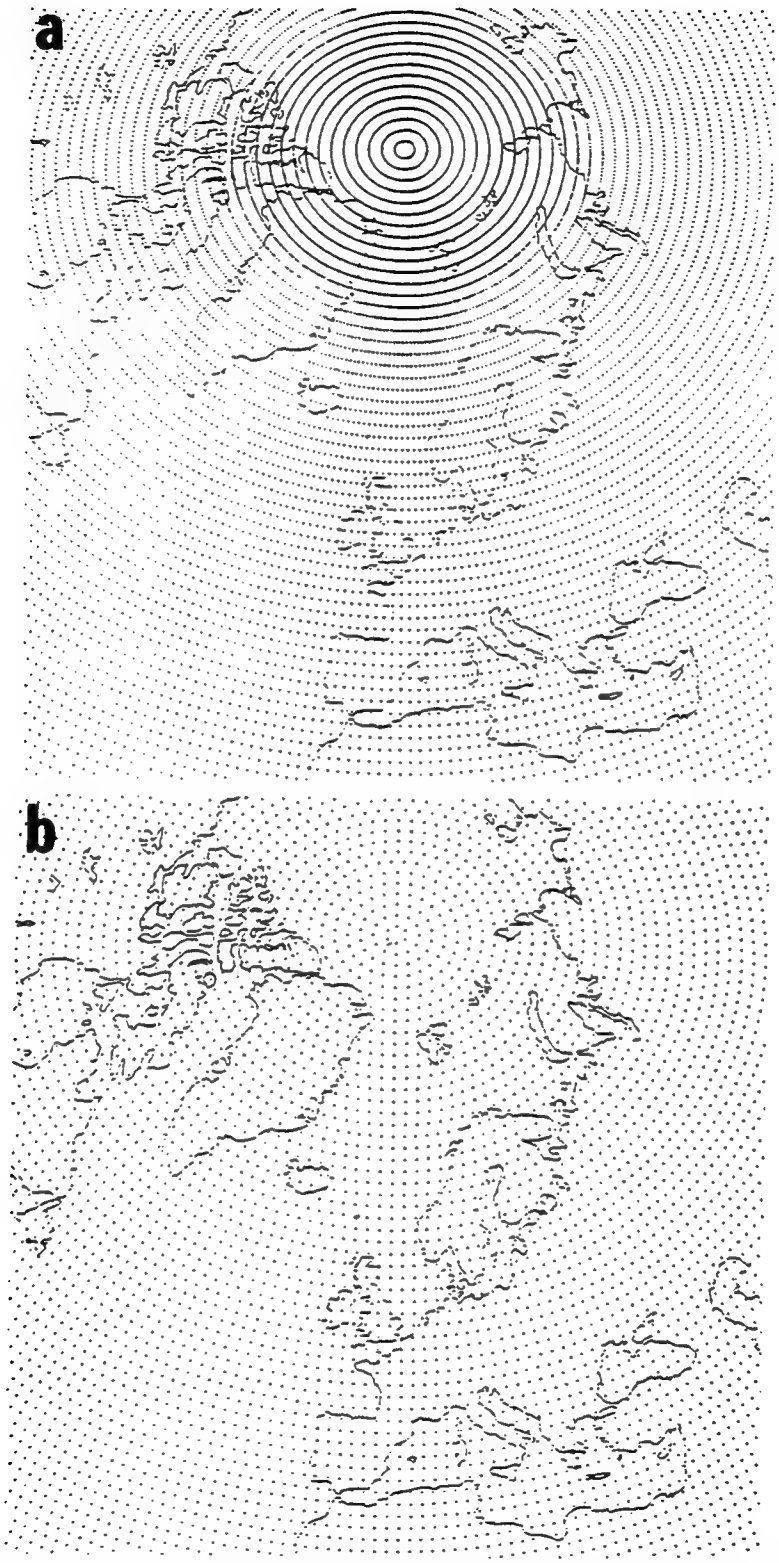

Figure 2. Comparison of Gridpoint Distribution Between (a) Full T106 NWP Model Gaussian Grid and (b) Equal-Area Grid

the meteorological analyses to determine the monthly mean altitude for each of these three σ levels at each transform gridpoint. This information was used to place each RTNEPH cloud report (with its altitude information) for the corresponding 3X3 array into the correct vertical cloud deck(s). In the transformed RTNEPH data sets, we cannot state where in the deck the clouds lie or how thick they are--only that they fall somewhere within the vertical domain of the deck. Given the variations of NWP model variables that can occur within limited extents of the vertical profiles, it is unfortunate that we cannot match the cloudiness with the corresponding vertical location in the model atmosphere. If the NWP model were very accurate in its vertical location of cloud-related weather phenomena, this inability to vertically match the observed cloud with the correct model layer would compromise the ability of the cloud diagnosis scheme. In practice, however, models exhibit vertical as well as horizontal random errors in the placement of weather phenomena. As the accuracy of NWP model forecasts and observed cloud depiction improves, we would expect diagnoses of cloudiness from NWP model forecasts to improve accordingly.

3.2 Development of Transform Grid RTNEPH Cloud Amount Data Sets

We obtained Northern Hemisphere (NH) RTNEPH cloud analysis data sets for January, February, July, and August 1991 from the U.S. Air Force Environmental Technical Applications Center (USAFETAC) in Asheville, NC on magnetic tapes. The data, as received from USAFETAC, were in climatological format, or "box-time" format. As shown in Figure 3, there are 60 RTNEPH "boxes" (boxes 1, 8, 57, and 64 lie completely off the hemisphere and are not included), each of which contains an array of 64X64 RTNEPH gridpoints for which a record of data is available. The analyses are available at 3-hour intervals. The box-time format means that all 3-hour periods in a given month for a given box are arranged on the tape contiguously before the data for the next box begins. For this project, we created files of the data in synoptic form--all 60 NH boxes in numerical order for the 0000, 0600, 1200, and 1800 UTC time periods of January and July 1991. In creating these files, we took note of

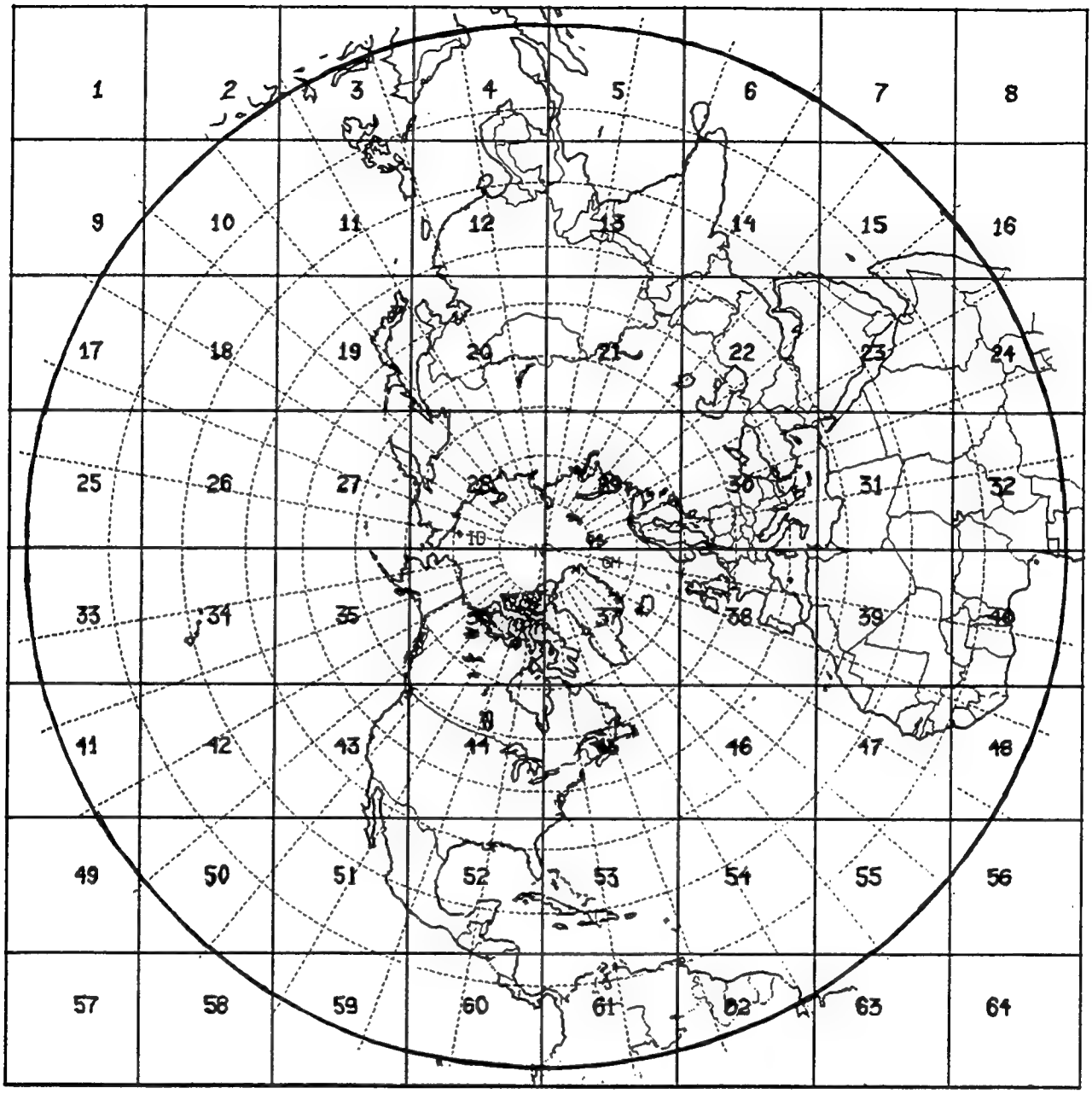

Figure 3. Northern Hemisphere RTNEPH Boxes Superimposed on the Polar-Stereographic Projection that Defines the RTNEPH One-Eighth Mesh Grid. Each RTNEPH Box is Made up of An Array of 64X64 Analysis Points. Data Are Available Only for the On-Hemisphere Analysis Points.

the indicator at each RTNEPH gridpoint that indicates the "age" of the cloudiness data--that is, how many hours ago the reported observation was made. For all RTNEPH gridpoints where the data was more than 2 hours old, we set a specially inserted flag (as the ninth data word) in the data record to indicate the data was untimely. For these points, we wrote only one record that included the original eight data words [indicating the box number, year, month, day, hour, "i" value, and "j" value (each of which now range 1-64), and a persistence indicator) and the ninth word, the timeliness flag. For all RTNEPH gridpoints with timely data (data no more than 2 hours old), we wrote this same nine-word record, but with the timeliness flag set to a different value when total cloud amount is zero than when it is nonzero. Then for timely, nonzero total cloud amount gridpoints only, we wrote a second record containing all of the original 28 words of cloud and weather information that was present in the RTNEPH data. By writing out the cloud and weather information only for timely, cloudy points, we were able to preserve all essential information on the RTNEPH cloud distribution while realizing a significant reduction in file storage requirements.

After developing the synoptic form of the RTNEPH data, we transformed each of the six-hour time periods for each month onto the transform grid. Horizontally, this involved locating the hemispheric RTNEPH grid index (1-512 in both "i" and "j" directions) of the RTNEPH gridpoint lying closest to each transform gridpoint. Once that RTNEPH point and its contiguous neighbors (a 3X3 array in all) were identified, each point was evaluated one at a time for its contribution (if any) to each of the three cloud decks of the transform grid (high, middle, low). For each layer (up to four) of cloudiness reported at the RTNEPH point, we took note of the base and top altitudes, the source flag (indicating what observing system was used at that point for that layer), and the cloud type. If a timely RTNEPH point had no contribution to a particular cloud deck, a contribution of zero cloud amount was made to an accumulating sum for that deck. If the RTNEPH point was untimely, it was considered missing and the number of points contributing to the 3X3 average was decremented by one. For cloudy, timely points, each cloud layer was considered in

turn and put into one of three categories according to the report cloud type: cumuloform (cumulonimbus and cumulus), stratocumulus (stratus and stratocumulus), and stratoform (all other types). Then the source flag was considered for the purpose of determining what information was used to place the cloud amount corresponding to that report layer in the appropriate deck. When a surface observation was indicated, the base altitude was used; when a satellite observation was indicated, the top altitude was used; when a radiosonde observation was indicated, both the base and top altitudes were used. If just the base or the top was used, the reported cloud amount was added to the accumulating sum in that deck into which the reported altitude lies. For cases where both base and top are indicated, any decks in which the base and top altitudes lie received that cloud amount as a contribution to their accumulating sums. Since it was possible for more than one reported cloud layer for a given RTNEPH point to fall within a certain cloud deck, we allowed only the layer with the maximum cloud amount to be contributed to that deck's accumulating sum. If a reported stratocumulus (by our categorization) layer lies in the middle cloud deck, we extended it into the low deck. We extended a reported cumuloform (by our categorization) layer to all decks below the deck that it fell within based on its altitude assignment. Any reported thin layer that lies in the high cloud deck was considered no cloud. After making these contributions to the accumulating sums for each of the timely points in the 3X3 array, we computed the average cloud amount in each deck by dividing the accumulated sum by the number of timely points, then rounding the result to the nearest 5%. We then flagged (for later use) all deck average cloud amounts made up mostly of cumuloform cloud reports. If the number of timely points was less than five, we set the deck average cloud amount to a flag indicating a missing cloud amount. Because timeliness was the same for all reported cloud layers at an RTNEPH point, all three decks have the same number of timely points. Thus, at any transform gridpoint, valid or missing average cloud amounts exist for all three decks. This means that over the hemisphere, exactly the same number of valid cloud amount transform gridpoints exist for all three decks.

Although the RTNEPH data set includes an estimate of total cloud cover at each gridpoint, we used it in this study only to determine if a given gridpoint was cloudy or not in forming the synoptic data set. The focus of our study is layered cloud amounts, from which we derived an estimate of total cloud amount for illustrative reasons only. We used the method of Trapnell[17] to estimate total cloud amount from any specification of deck cloud amounts that we had. This two-step method uses low and middle fractional cloud amount (CA) to arrive at a preliminary value of total fractional cloud amount (TC):

$$TC_{prelim} = MAX\ (CA_m,\ CA_l) + [1 - MAX\ (CA_m,\ CA_l)] \cdot MIN(CA_m, CA_l) \cdot \alpha \qquad (1)$$

which is then used with high cloud amount to arrive at a final value of total cloud amount:

$$TC = MAX\ (TC_{prelim},\ CA_h) + [1 - MAX\ (TC_{prelim},\ CA_h)] \cdot MIN(TC_{prelim},\ CA_h) \cdot \beta \qquad (2)$$

where α and β are parameters that allow the assumed amount of cloud overlap to be varied. We followed Trapnell's tuned overlap assumption and set $\alpha = \beta = 0.55$. Again, it should be emphasized that total cloud estimates derived in this manner were produced only from deck cloud amount specifications (either transformed RTNEPH or diagnosed from forecasts). Cianciolo[22] used the reported RTNEPH total cloud amounts to develop regression-based predictor-predictand relationships. In this study, we chose to develop such relationships separately for each cloud deck, and not for total cloud. Total cloud relationships may be derived and applied in future studies.

3.3 Evaluation of the Transform Grid RTNEPH Cloud Amount Data Sets

We sought independent cloudiness data samples as a check on the veracity of our cloud amount representations on the transform grid. Because of a dearth of quantitative independent data, we made only qualitative comparisons of our cloud amount distributions with outside data sources. Our first comparisons were with Meteosat satellite imagery depicted in the Deutscher Wetterdienst daily weather

bulletin. Figure 4 shows the infrared satellite imagery on 0000 UTC 19 January 1991 for a region of the northeastern Atlantic Ocean and western Europe. Two distinct cloud masses are apparent in the imagery. What is not apparent in the photo reproduction of the imagery is the region of scattered cloudiness in the northern part of the region between the two cloud masses, and the clear area in the region immediately west of the more eastward cloud mass, just where it bends westward. Also, it is difficult to tell from the reproduction what may constitute high, middle, or low clouds. For this reason, it is not possible to rigorously verify the transformed RTNEPH cloud amount representations included in the figure. The position of the maxima and minima seem consistent with the imagery. The well-defined frontal band, apparent especially in the middle deck, is in a position consistent with that depicted in the imagery. Also, the minimum (less than 20 percent) region just behind the eastward frontal band, just NW of the Iberian peninsula, is correctly positioned. Our sense is that perhaps the RTNEPH high cloud amounts may be underestimating the actual amounts of high cloud, especially in the central portion of the eastward frontal band. However, it is not possible to state this with certainty.

In Figure 5, we show a summertime comparison with imagery over the same region, for the date 0000 UTC 23 July 1991. We felt that brighter amounts of the major cloud masses would probably contain more cloudiness than what is depicted on the RTNEPH high cloud figure. This apparent underestimate may be due to either underestimation of the altitude of cirrus clouds in the RTNEPH data, our selection of the NWP model σ level defining the high cloud deck base (which is on average about 400 m higher in July than in January for the mid-latitudes), our decision to ignore thin high clouds, or a combination of the three. This may suggest the use of a temperature criterion for setting the high cloud deck base in future work. In any case, we feel that the pattern of cloudiness, represented by the middle and low clouds, is consistent with the imagery over Europe and the Atlantic. Notice, for example, the relatively middle cloud-free region SW of Iceland, but where low clouds exist. This coincides with the "darker" (lower) cloud images westward of the frontal cloud mass extending southward from Iceland in the imagery. Note that the low and middle deck

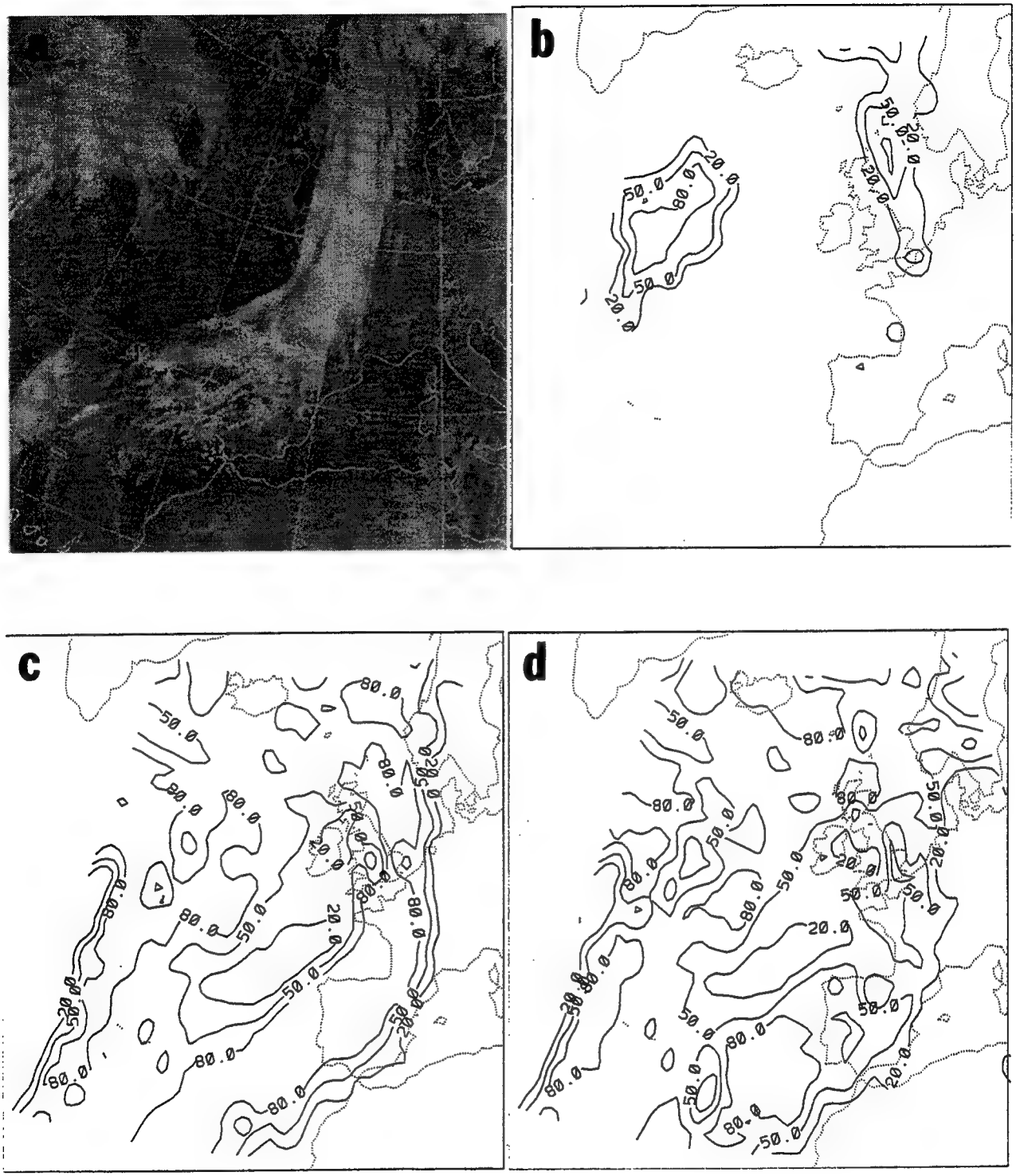

Figure 4. Cloud Distribution on 0000 UTC 19 January 1991 as Depicted by (a) Meteosat Satellite Imagery and RTNEPH Transform Grid Cloud Amounts (in Percent) in the (b) High, (c) Middle, and (d) Low Cloud Decks.

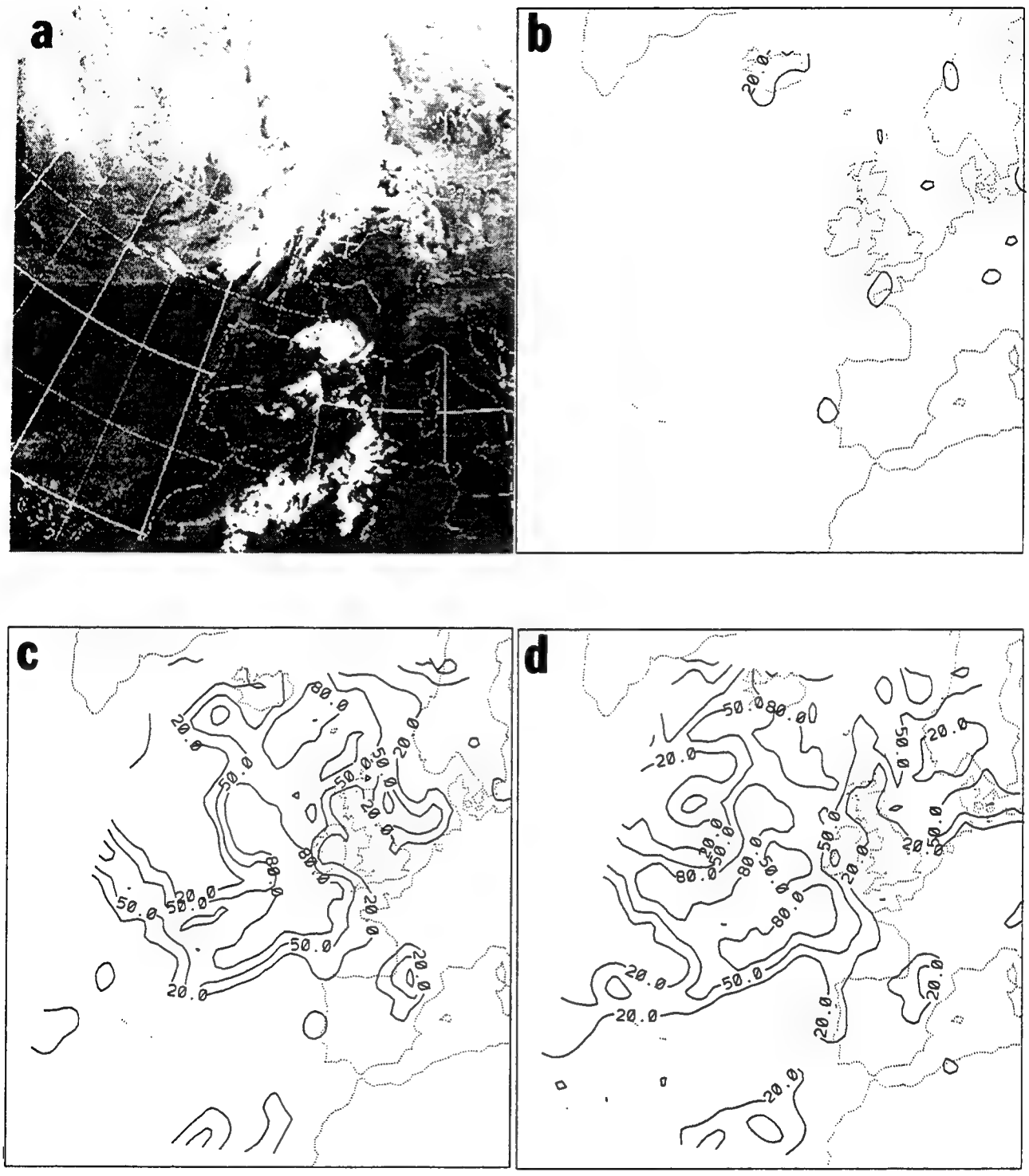

Figure 5. Same as in Figure 4 for 0000 UTC 23 July 1991.

RTNEPH includes the isolated convective cloudiness over the Pyrenees Mountains and the low cloud west of North Africa. However, it missed the brighter (higher) clouds located over North Africa and extending into the Mediterranean Sea.

In addition to our comparisons with Meteosat imagery, we also computed global and zonal average statistics and compared them with previously published statistics of 3DNEPH. Table 2 documents the comparison of our 1991 monthly, hemispheric averaged transformed RTNEPH cloud amounts with those published by Henderson-Sellers[29] for 1979 3DNEPH statistics. In the latter study, cloud decks were defined as: low cloud--0 to 6500 feet above ground level; middle cloud--6500 to 20,000 feet above ground level; high cloud--above 20,000 feet above ground level. Over the ocean and non-mountainous land, the base of the middle cloud deck chosen in the 3DNEPH study would be 200-300 m lower than in our study, and the base of the high cloud deck would be 500-900 m lower than in our study. Thus, we would expect more high cloud in the 3DNEPH study; the effect of the middle cloud deck base difference is uncertain. Also, recall that only timely data were used in the current study, whereas all 3DNEPH data were used in the 1979 study.

Table 2. NH Monthly Average Cloud Amounts (Percent)

	January		July	
	3DNEPH 1979	RTNEPH 1991	3DNEPH 1979	RTNEPH 1991
High	10	5	13	5
Middle	27	37	32	43
Low	33	38	44	42

Both studies show an increasing trend in middle and low cloudiness from winter to summer. Here again, because of the differences in years of the study and study designs, our comparisons can only be qualitative.

We sought to validate the latitudinal variations of the zonal averages of the transformed RTNEPH cloud amounts. Here again, we compared our results with the 1979 3DNEPH study.[29] In Figures 6 and 7, we show the zonal average cloud amounts

for the two studies for January and July respectively. The two sets of curves for January show the same general trends--a cloud cover minimum at 10-20N, a mid-latitude maximum, then in middle and total cloud a relative minimum at 70N and an increase poleward, while low and high clouds decrease in amount toward the Pole. The crossover point (about 50N) where middle cloud exceeds low cloud agrees also, although low-middle cloud differences appear smaller in the 1991 RTNEPH statistics equatorward of this point. The apparent underspecification of high cloud in the 1991 transformed RTNEPH is more striking in the July zonal average cloud amount comparison. The two peaks, at 10N and 50-60N, are much lower in the 1991 RTNEPH than in the 1979 3DNEPH. Perhaps as a consequence, the July 1991 RTNEPH middle cloud amounts exceed their July 1979 3DNEPH counterparts at 10N by about 30 percent. At 50-60N, low and middle clouds from the two studies are in agreement. However, in the two 3DNEPH minima regions--30N and 90N--the middle clouds (and the low clouds at 90N) exhibit greater zonal averages. Qualitatively, we can say that the two data sets agree in latitudinal trends. Their absolute differences cannot be resolved definitively because of the study year and design differences.

In Table 3, we present transformed 1991 RTNEPH cloud amount statistics for January and July. In our computations, we partitioned the hemisphere two ways--30° latitude bands, and tropical (0-30N) and extratropical (30-90N) water and land regions. For our purposes, transform gridpoints representing areas less than half water covered were considered land points.

We used this partitioning to determine whether or not the statistics for the various regions would reveal significant differences that might suggest the benefit of regional applications of the forecasting methods. The greatest regional variations of the cloud amount average and standard deviation statistics were found in the tropical and extratropical land and water stratification. For this reason, we present only the statistics for this stratification of regions, plus the hemispheric statistics.

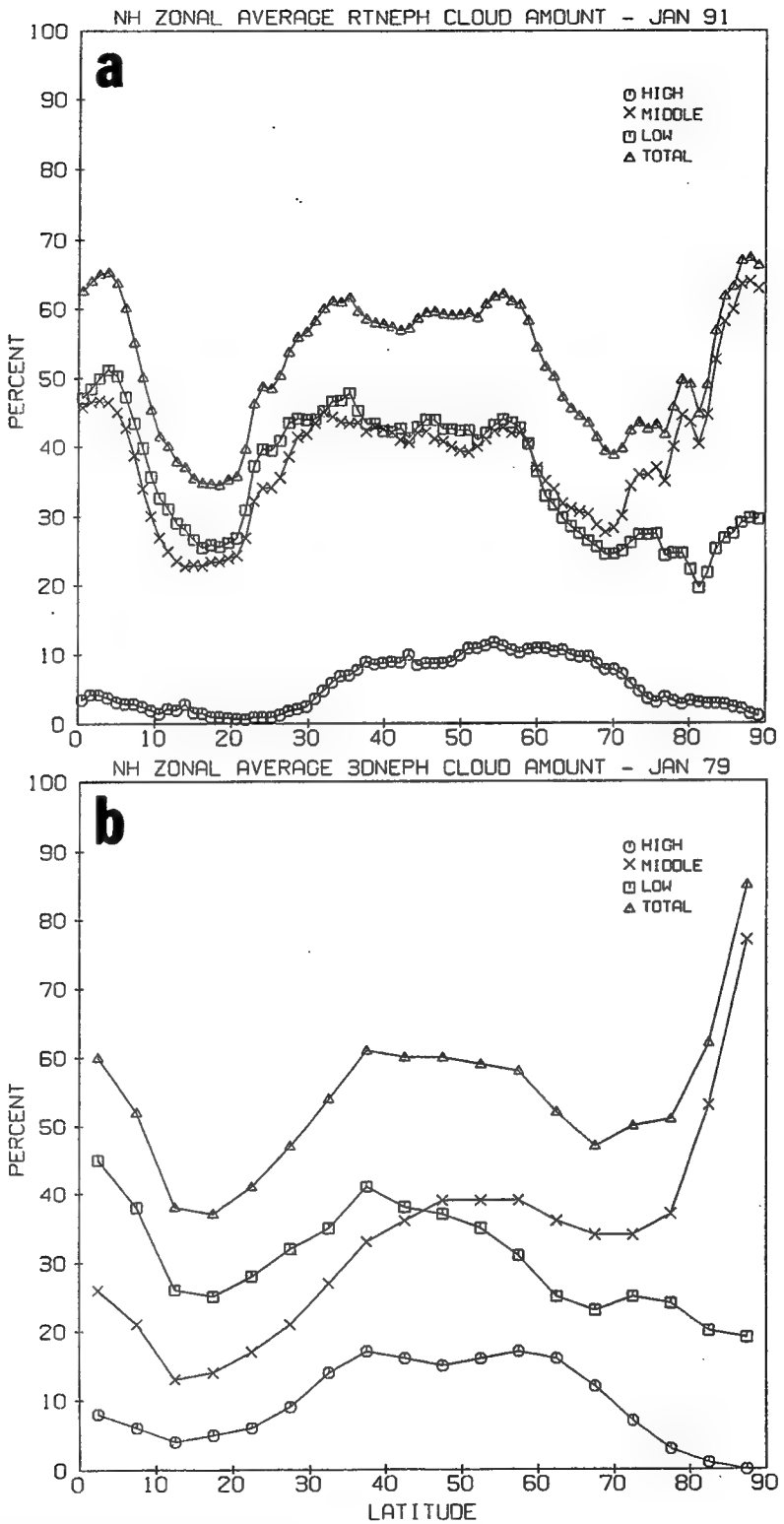

Figure 6. **Zonal, Monthly Average Northern Hemisphere Cloud Amounts (Percent) for 1979 3DNEPH and 1991 RTNEPH for January.**

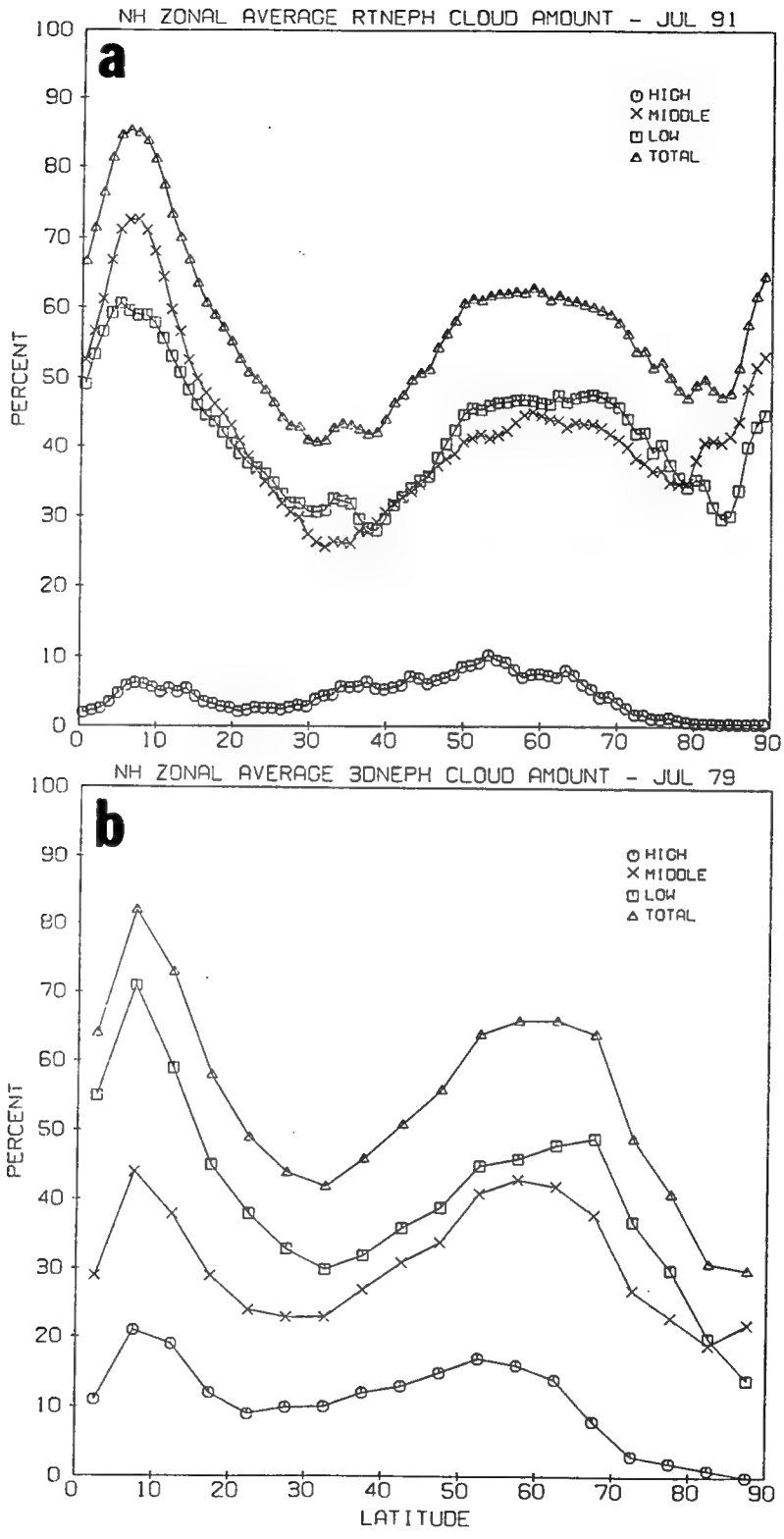

Figure 7. Same as in Figure 6 for July.

Table 3. NH Transformed RTNEPH Cloud Amount Statistics

January 1991

Average Cloud Amount Percent

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	2.3	1.9	9.7	6.7	5.4
Middle	22.0	38.1	27.3	54.7	36.9
Low	23.0	43.5	27.5	51.0	37.9
Total	33.1	54.5	44.8	67.4	52.0

Cloud Amount Standard Deviation Percent

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	10.1	10.4	20.5	19.1	16.5
Middle	32.9	38.5	33.8	37.6	38.1
Low	30.4	35.3	34.0	37.1	36.4
Total	37.4	39.0	37.6	36.2	39.3

July 1991

Average Cloud Amount Percent

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	4.1	3.6	8.9	2.4	4.8
Middle	45.8	52.7	31.6	42.1	43.1
Low	36.8	50.3	37.2	41.0	42.5
Total	56.2	66.9	49.7	56.6	58.0

Cloud Amount Standard Deviation Percent

Cloud Deck	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
High	14.0	14.7	18.1	9.6	14.8
Middle	42.0	40.9	34.8	38.0	39.5
Low	34.4	35.6	34.3	34.0	35.1
Total	41.7	37.7	37.2	37.6	38.7

Table 4 lists the six-hour time period average number and percentage of equal area gridpoint values available in each region for each month.

Table 4. Number and Percentage of Gridpoint Cloud Amount Values Available in Each Region

Six-Hour Time Average Values

Region	Total Gridpoints	% of Hemis	Jan % Avail	Jul % Avail
0-30N Land	2354	14	52	47
0-30N Water	5899	36	49	50
30-90N Land	3971	24	65	63
30-90N Water	4155	25	53	58
0-90N	16379	100	54	55

Looking first at the geographic stratification as given in Table 4, we see that roughly equal areas of the hemisphere lie poleward and equatorward of 30N as expected. The breakdown between land points and water points is about equal poleward of 30N, but in the tropics, the surface area of water is slightly more than 2.5 times that of land. The extratropical regions have a greater number of gridpoints with non-missing cloud amounts because of the fact that polar orbiting satellites, upon which the RTNEPH primarily depends, give overlapping surveillance in the extratropics, while leaving gaps in the coverage in the tropics. The latter fact is one of the primary reasons why geostationary imagery will be included in the cloud analyses produced in the CDFS-II era. However, even in the tropical land region in July, over a 10-day period of twice-daily forecasts, the development sample would contain more than 20,000 data values at each forecast time. In our study, time did not allow us to develop separate predictive equations for each of these regions. We were only able to develop statistical predictor-predictand relationships for the entire hemisphere. However, it appears from Table 4 that it would be possible to develop stable statistical relationships for each region.

Table 3 suggests that there may be some benefit from regional statistical predictive relationships. Looking at the average cloud amounts, we see a clear

tendency for greater middle and low cloudiness over water in both tropical and extratropical regimes, especially in January. In the high cloud, we see just the opposite trend--more cloudiness over land than over water surfaces. Given the significantly smaller amounts of high cloud than middle and low cloud, and the fact that the high cloud standard deviation is so much larger than the average, the water-land differences may be less significant in the high deck. In January, the extratropical regions exhibit more cloudiness than their tropical counterparts in all three decks. In July, this trend reverses except for high and low cloud averages over land. Generally, the larger standard deviations are associated with the larger averages, but they are not proportional. For example, in January, the middle and low averages over extratropical water are more than 2 times greater than over tropical land, but the standard deviations are only about 20 percent greater. Since statistical methods tend to forecast better in regions where temporal and spatial variations are smaller, we might expect our schemes to perform only modestly better in January forecasts over tropical land than over extratropical water surfaces.

In predicting our ability to forecast cloud amount using statistical methods, it is also useful to consider the frequency distribution of the various cloud amount categories. As stated previously, we maintained the partitioning of cloud amount into 5 percent categories in the transformation of the RTNEPH data onto the equal-area grid. We therefore can plot histograms of the frequency of occurrence of clouds in each of the decks for each region of stratification. We display these plots in Figure 8 for the entire hemisphere (the major features of the regional plots are similar). The most striking feature of these histograms is the large frequency of occurrence of 0 percent cloud amount (clear). The high cloud deck is clear at 80 percent of the gridpoint-times in both months. The entire column is clear over 20 percent of the gridpoint-times in January and 15 percent in July. The probability of decks combining for 100 percent cloud cover (overcast) is 18 percent for January and 22 percent in July. In the three cloud decks, the frequency of occurrence remains very steady between 5 percent and 95 percent inclusive. According to Glahn,[23] linear schemes tend to work best with data when population distribution is normal in both predictor

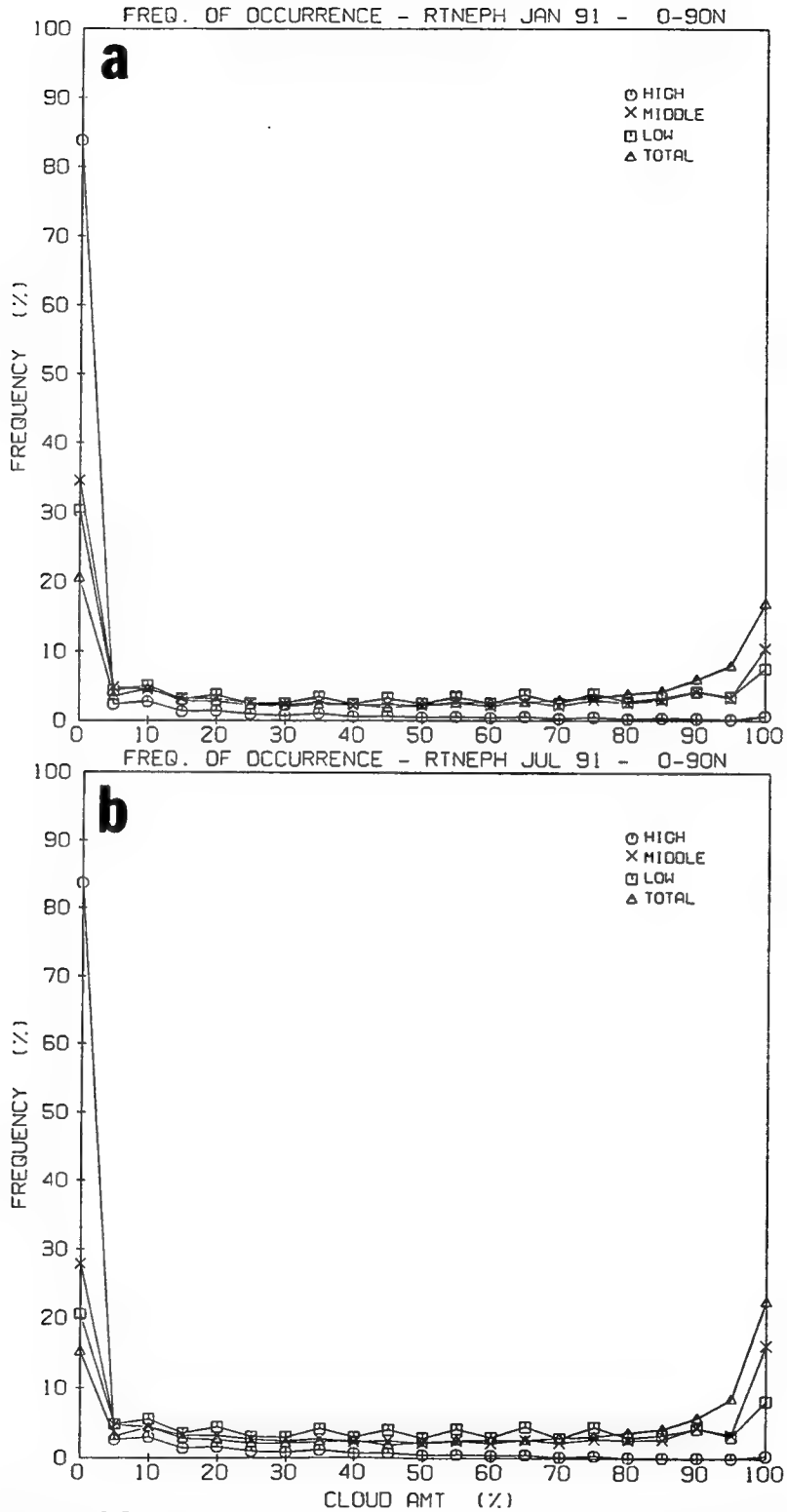

Figure 8. **Frequency of Occurrence (Percentage of All Transform Gridpoints for Four Times Per Day Over the Month) of Cloud Amount in the Three Decks and Total Cloud for (a) January and (b) July 1991.**

and predictand. Here, we see an equal spread of the predictand with spikes at either end of the spectrum. As we shall see, the multiple linear regression scheme had to be modified to account for this bimodal distripution.

3.4 Development of the NWP Model Forecast Data Sets

We chose to use the Phillips Laboratory Global Spectral Model (PL GSM) as the NWP model responsible for making the weather forecasts from which we would develop and apply statistical forecast procedures. The version of the PL GSM used in this study was PL-92. This version uses the same numerical approximations and physical parameterizations as does the version described by Norquist et al.[30], PL-91. However, PL-92 includes the generalized truncation formulation of the numerics and incorporates the model optimization upgrades described by Nehrkorn et al.[31] The generalized truncation formulation allows us to run the model in rhomboidal or triangular spectral truncation mode. The optimization upgrades allow it to execute more efficiently because of code vectorization, and in parallel on several processors simultaneously because of multitasking directives included in the code. The PL GSM was developed specifically for cloud forecasting research. This study serves as an attempt to determine a baseline capability for the NWP model-cloud diagnostic approach to cloud forecasting. Any future improvements that we make in the forecasting skill of the model can then be evaluated for their impact on the accuracy of cloud forecasts produced by diagnosing cloud cover from the model forecasts.

To initialize the PL GSM forecasts, we acquired global meteorological analysis data sets from the European Centre for Medium Range Weather Forecasts (ECMWF). We obtained a subset of the ECMWF/TOGA Advanced Operational Analysis Data Sets, namely the 0000 and 1200 UTC global analyses for the entire months of January, February, July, and August 1991. These analyses were in spherical harmonic representation of a triangular spectral truncation of 106 waves (T106). The variables included in the analysis set were geopotential (which we converted to geopotential height, Z), temperature (T), vertical velocity (which we did not use),

zonal wind component (u), meridional wind component (v) and relative humidity (RH). These data were available in the analyses on the following pressure surfaces: 1000, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, and 10 hPa.

The steps involved in preparing the initial conditions were as follows. First, the T106 version of the model surface terrain was constructed. Using a U.S. Navy 10' terrain data base, we identified all 10' terrain gridpoints that fell within each gridbox of our 320 longitude X 160 Gaussian latitude PL GSM computational grid (corresponding to T106). We then averaged the 10' terrain values in each computational gridbox to arrive at the mean orography used in our PL GSM forecasts. We executed a forward and backward spectral transform on these 320 X 160 terrain values to effectively truncate the spectral resolution to T106. Next, we unpacked the T106 spectral harmonics analysis values of Z, T, u, v, and RH onto the 320 X 160 computational grid on each of the 14 pressure surfaces. The next step was to enter our pre-processor which carried out the following steps: (1) use a hydrostatic relationship [geopotential as a quadratic function of the natural logarithm (ln) of pressure (p)] to deduce a model terrain surface pressure (p_s) at each computational gridpoint, (2) interpolate T, u, v, and RH linearly in ln p from pressure surfaces to 22 model σ layers (approximately the midpoints between the σ levels, positioned at $\Delta\sigma = 0.05$ from $\sigma = 0$ to $\sigma = 0.95$, then $\Delta\sigma = 0.025, 0.015, 0.010$ below $\sigma = 0.95$) at each computational gridpoint, (3) convert RH to specific humidity (q) using the method of Derickson and Cotton[32] (an efficient algorithmic form of the Murray[33] formulation for the calculation of saturation vapor pressure) at each computational gridpoint, and (4) perform a forward spectral transform on u, v (converted to vorticity and divergence), T, q, and ln p_s. This representation of the analysis is referred to as T106L22, where L22 signifies the 22 σ layers in the vertical structure of the model atmosphere. Finally, we applied two iterations of a nonlinear normal mode initialization using five vertical modes where, for each iteration, the PL GSM provided full model (including diabatic term) tendencies to the initialization procedure. These initialized analyses, constructed for 0000 and 1200 UTC analysis for the first 24 days of January and July 1991, constituted our initial conditions.

Before this study, we had not yet executed the PL GSM at the T106L22 resolution. Earlier experimentation with the PL-91 version was executed at R30L18 (R represents rhomboidal truncation). In the process of converting PL-91 to PL-92, we ran both versions of the model at R30L18 using identical initial conditions until we were satisfied that both versions were yielding essentially the same forecasts. We then executed the T106L22 version of PL-92 and compared the resulting forecasts with the R30L18 forecasts of PL-92. Figures 9 and 10 show such a comparison for the 48-hour forecasts of 500 hPa Z and mean sea level pressure (pmsl) initialized at 0000 UTC 1 January 1991. Also included in the figures are the ECMWF analyses for the verification time, 0000 UTC 3 January 1991. Note that the forecast fields position the major synoptic feature correctly, but lack the intensity indicated in the analyses. The lack of intensity can be attributed at least in part to the initialization process, a fact we noticed by comparing maps of ECMWF analysis 500 hPa Z and pmsl with the corresponding initialized fields. In fact, we noticed that some synoptic features of the 48-hour forecasts were closer in intensity to the verifying analyses than were the 0-hour forecasts to their corresponding features. For this reason, we verified maps using the uninitialized analyses, but for all other forms of verification, we used the initialized analyses to isolate the forecast error. For example, in computing the anomaly correlation for the NH north of 20N, we used the initialized analyses as verification for both the R30L18 and T106L22 versions of PL-92. The 12-hour forecasts initialized 0000 UTC 1 January 1991 had anomaly correlations of 0.987 and 0.985 respectively, and both versions had 48-hour anomaly correlations of 0.923. In this project, we made no attempt to "tune" the model when we ran it at the higher resolution. In spite of this, we found that the forecast fields from the two resolutions, when evaluated against the initialized analyses, were very similar. We concluded from this fact that, for this project, it would not be necessary to tune the T106L22 version of PL-92 to yield reasonable NWP model forecasts for cloud diagnosis. Hereafter in this report, any reference to PL GSM forecasts will refer to the use of the PL-92 version unless otherwise specified.

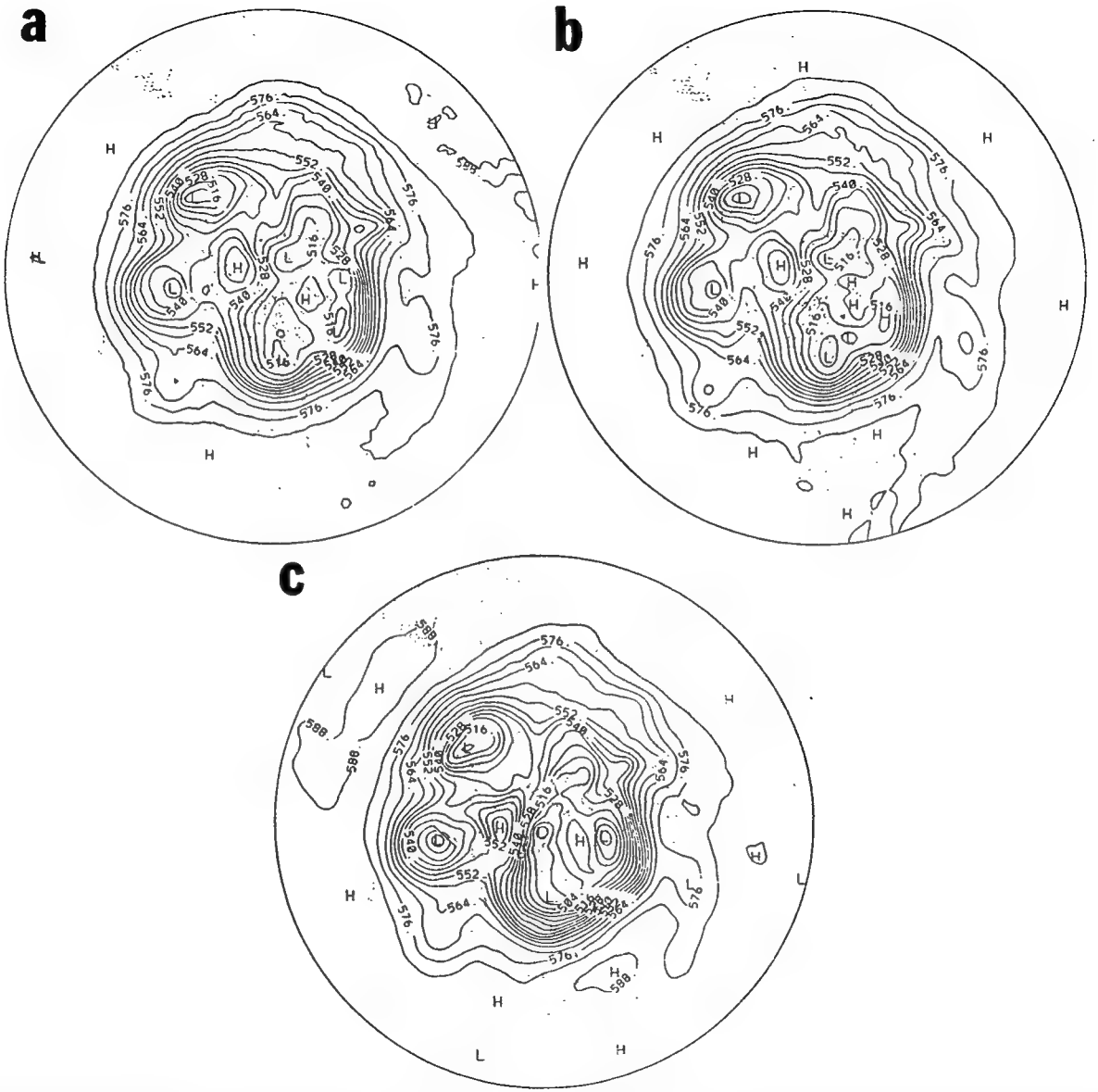

Figure 9. Northern Hemisphere 500 hPa Geopotential Height (decameters) Valid 0000 UTC 3 January 1991 from (a) 48-Hour T106L22 PL-92 Forecast, (b) 48-Hour R30L18 PL-92 Forecast, (c) ECMWF Analysis.

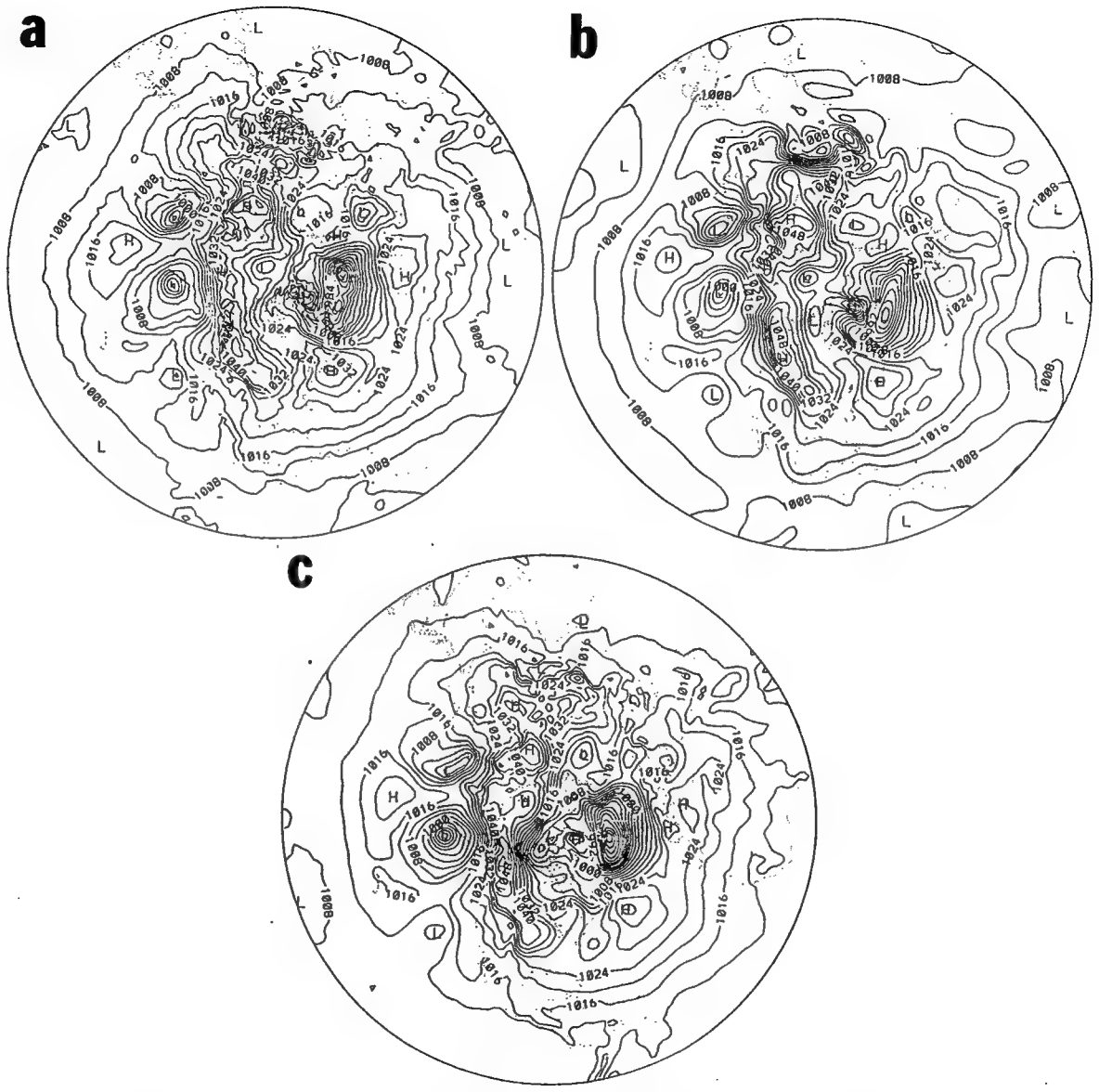

Figure 10. Same as in Figure 9 for Mean Sea Level Pressure (hPa).

We have displayed the zonal cross sections of u, T, and RH RMSE and mean error (bias) for the 12-hour and 48-hour T106L22 PL GSM forecasts initialized 0000 UTC 15 January 1991 and 15 July 1991 in Figures 11-16. These can be compared for reference with similar zonal cross sections of PL-91 forecasts shown by Norquist et al.[30] In the present set of cross sections, we have added the RMSE of RH because of our expectation that accuracy of RH forecasts will significantly influence cloud distribution diagnosed from the forecasts. In both months, the RH RMSE increases between forecast hours 12 and 48, suggesting that the spatial correlation with observed cloudiness may be weaker at 48 hours. We expect that this degradation of RH forecast accuracy with forecast time may lead to a similar degradation in the ability to forecast cloudiness at longer forecast times using a diagnostic method.

We executed 48-hour T106L22 PL GSM forecasts initialized at 0000 and 1200 UTC for the periods 1-24 January and 1-24 July 1991. Although we saved the forecast fields at 6-hour intervals of forecast time in anticipation of 6-hour diagnoses of cloud amount, time allowed us to develop and apply cloud diagnostic procedures only at forecast hours 12 and 48. For each 6-hour interval forecast time, we computed the transform grid values of deck average RH (RH_{ave}), deck maximum RH (RH_{max}), and the multi-layer and surface NWP variables listed in Table 1. Except where otherwise specified, the multi-layer NWP values represent the deck average of the quantity. In our case, this is the average of five σ layers in the low deck ($0.80<\sigma<0.99$), seven σ layers in the middle deck ($0.45<\sigma<0.80$), and 3-5 σ layers in the high deck ($\sigma_{lat}<\sigma<0.45$, where $\sigma_{lat} = 0.20$ for lat = 0-20N, $\sigma_{lat} = 0.25$ for lat = 20-65N, and $\sigma_{lat} = 0.30$ for lat = 65-90N). The RH_{ave} and RH_{max} values were used in the development of the CCA curves over a moving 7-day development period. Data sets of these values were constructed from forecasts over the entire 24-day period in each month. The multi-layer and surface NWP variable data sets were constructed from forecasts initialized from days 8-24 of each month. These were used in a moving 10-day development period to derive the statistical predictor-predictand relationships using the MLR procedure. Starting the CCA development 7 days before the MLR development ensured that CCA cloud amount forecasts (resulting from applying the

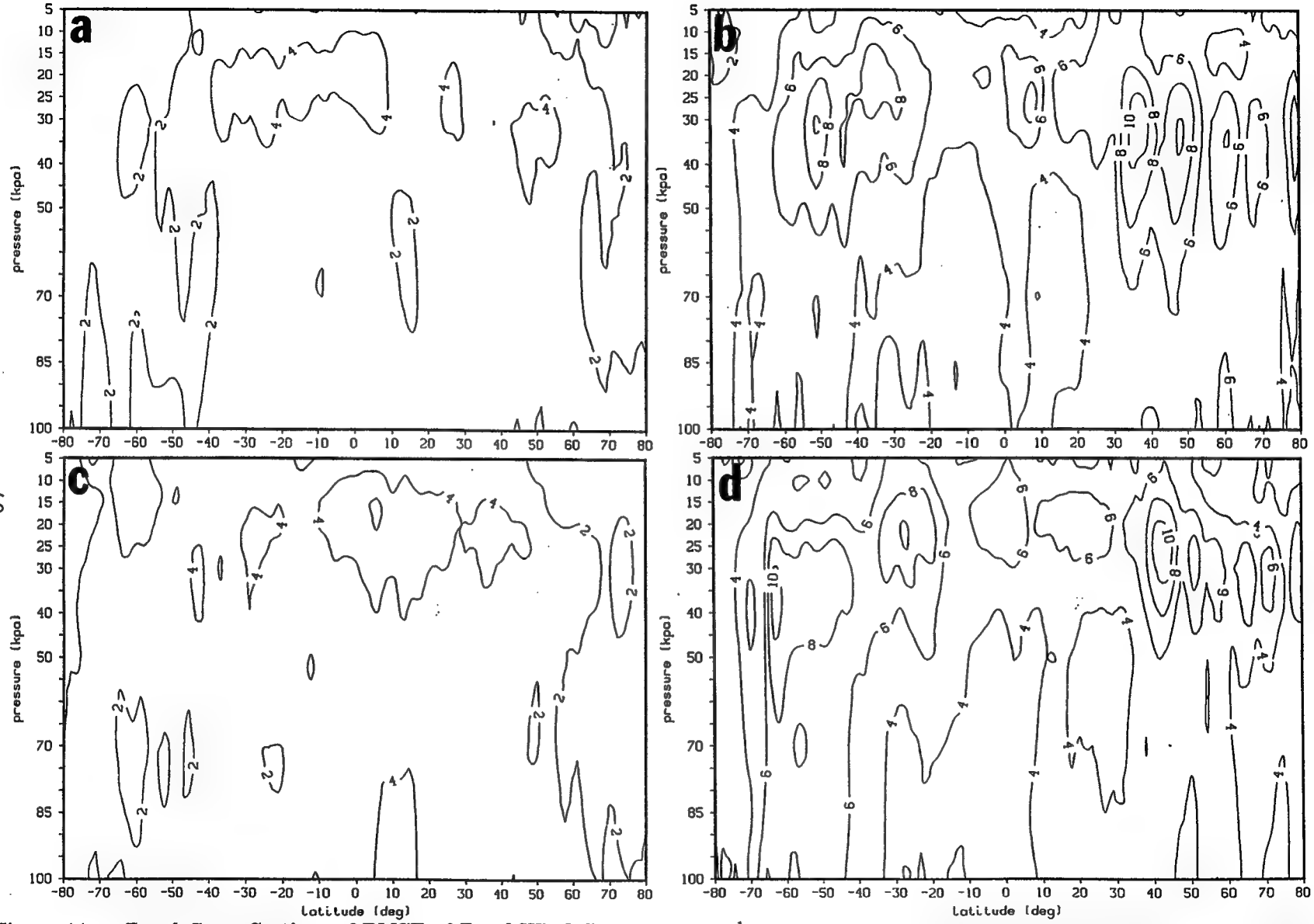

Figure 11. Zonal Cross Sections of RMSE of Zonal Wind Component (ms^{-1}) Computed on Indicated Pressure Surfaces at 1.125 Degree Latitude Intervals for T106L22 PL GSM (a) 12-Hour and (b) 48-Hour Forecasts Initialized 0000 UTC 15 January 1991, and (c) 12-Hour and (d) 48-Hour Forecasts Initialized 0000 UTC 15 July 1991.

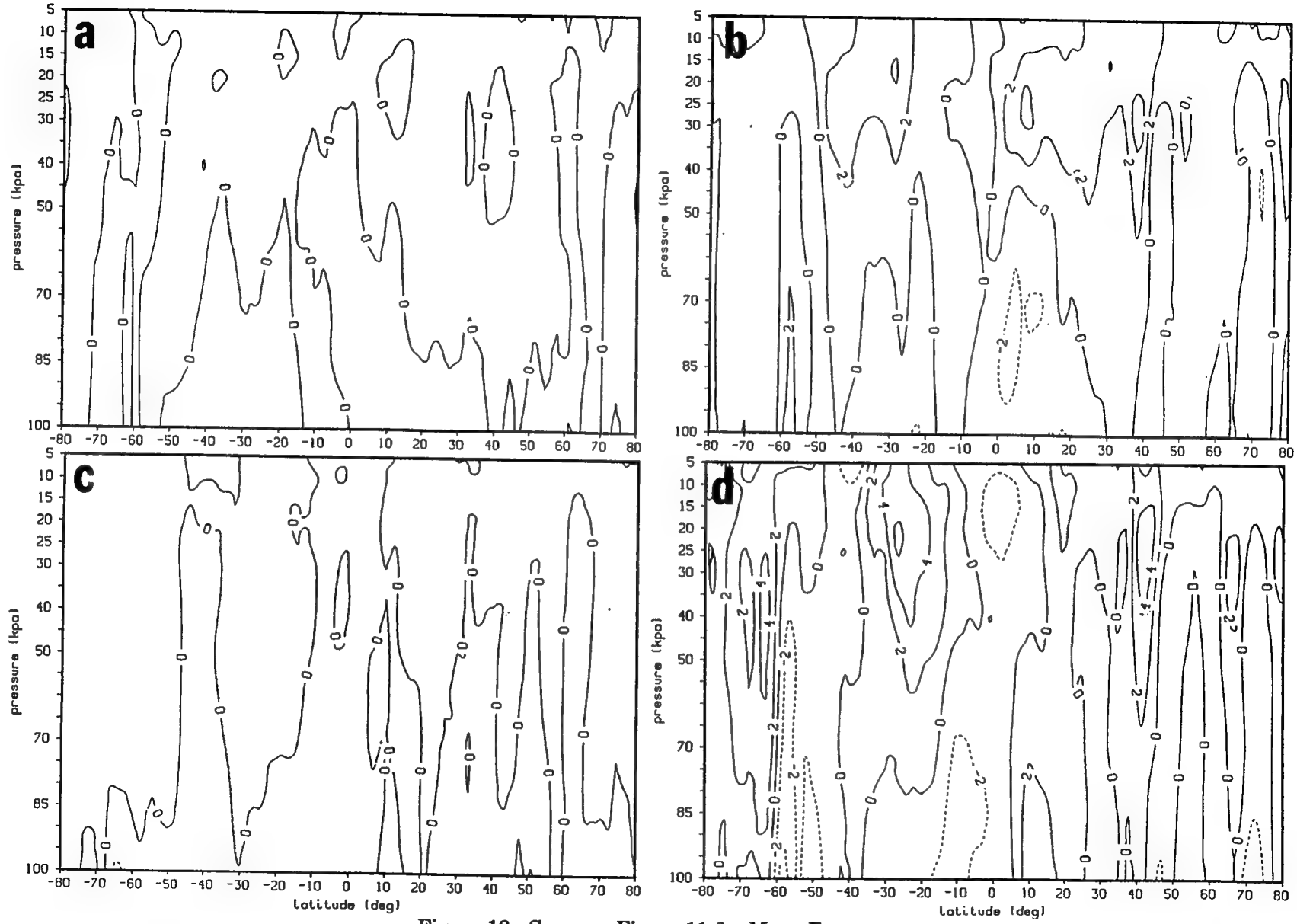

Figure 12. Same as Figure 11 for Mean Error.

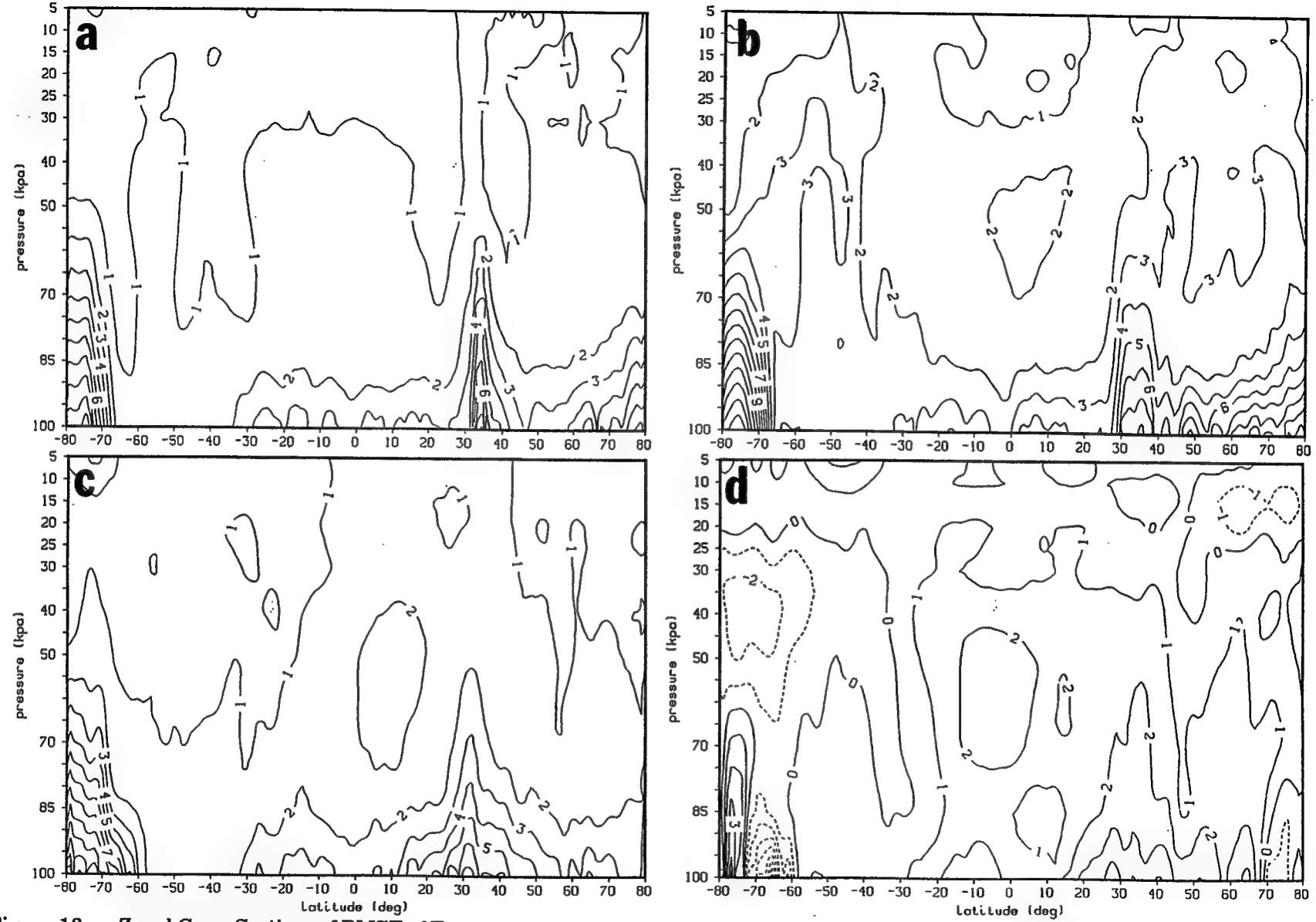

Figure 13. Zonal Cross Sections of RMSE of Temperature (K) Computed on Indicated Pressure Surfaces at 1.125 Degree Latitude Intervals for T106L22 PL GSM (a) 12-Hour and (b) 48-Hour Forecasts Initialized 0000 UTC 15 January 1991, and (c) 12-Hour and (d) 48-Hour Forecasts Initialized 0000 UTC 15 July 1991.

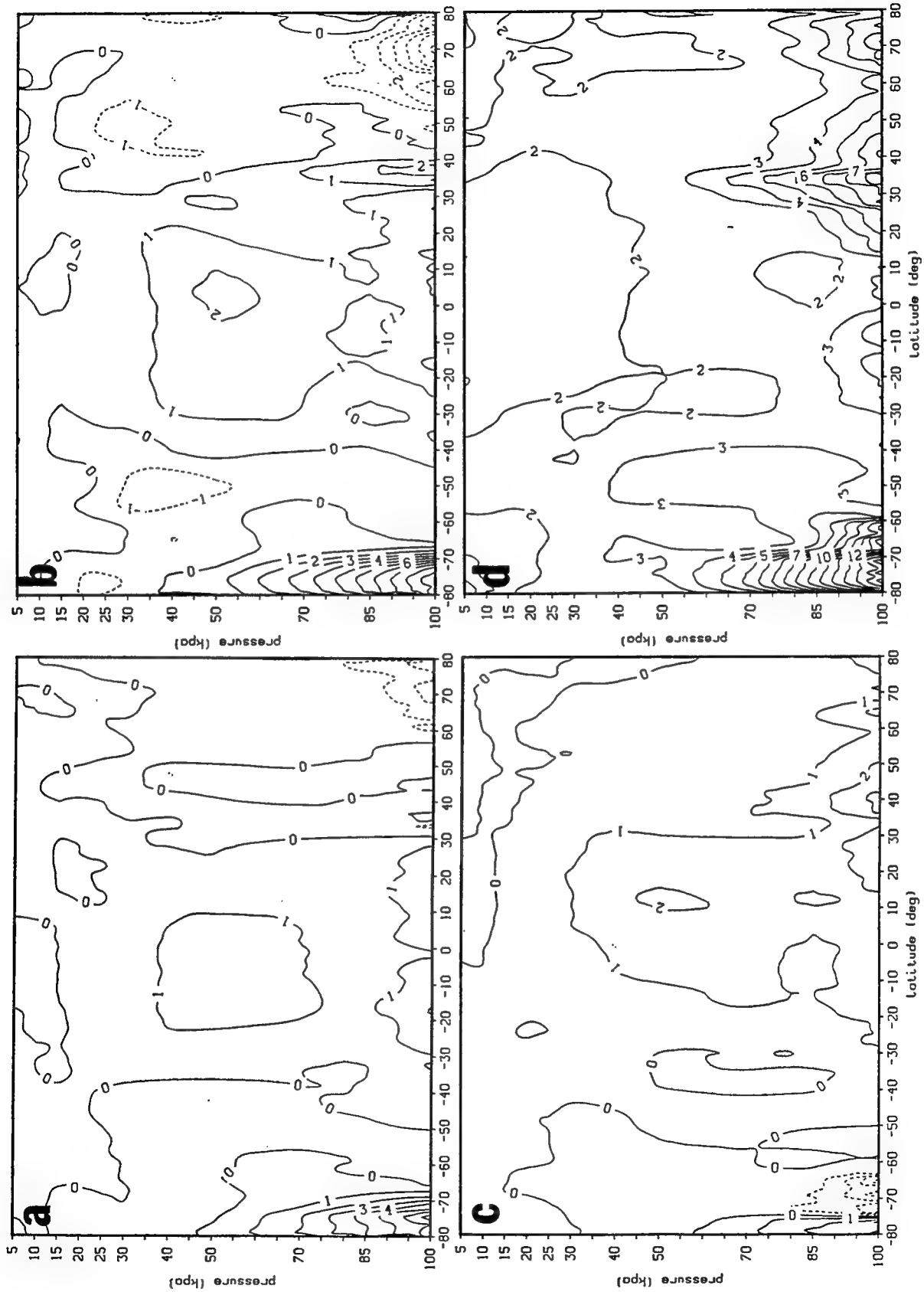

Figure 14. Same as Figure 13 for Mean Error.

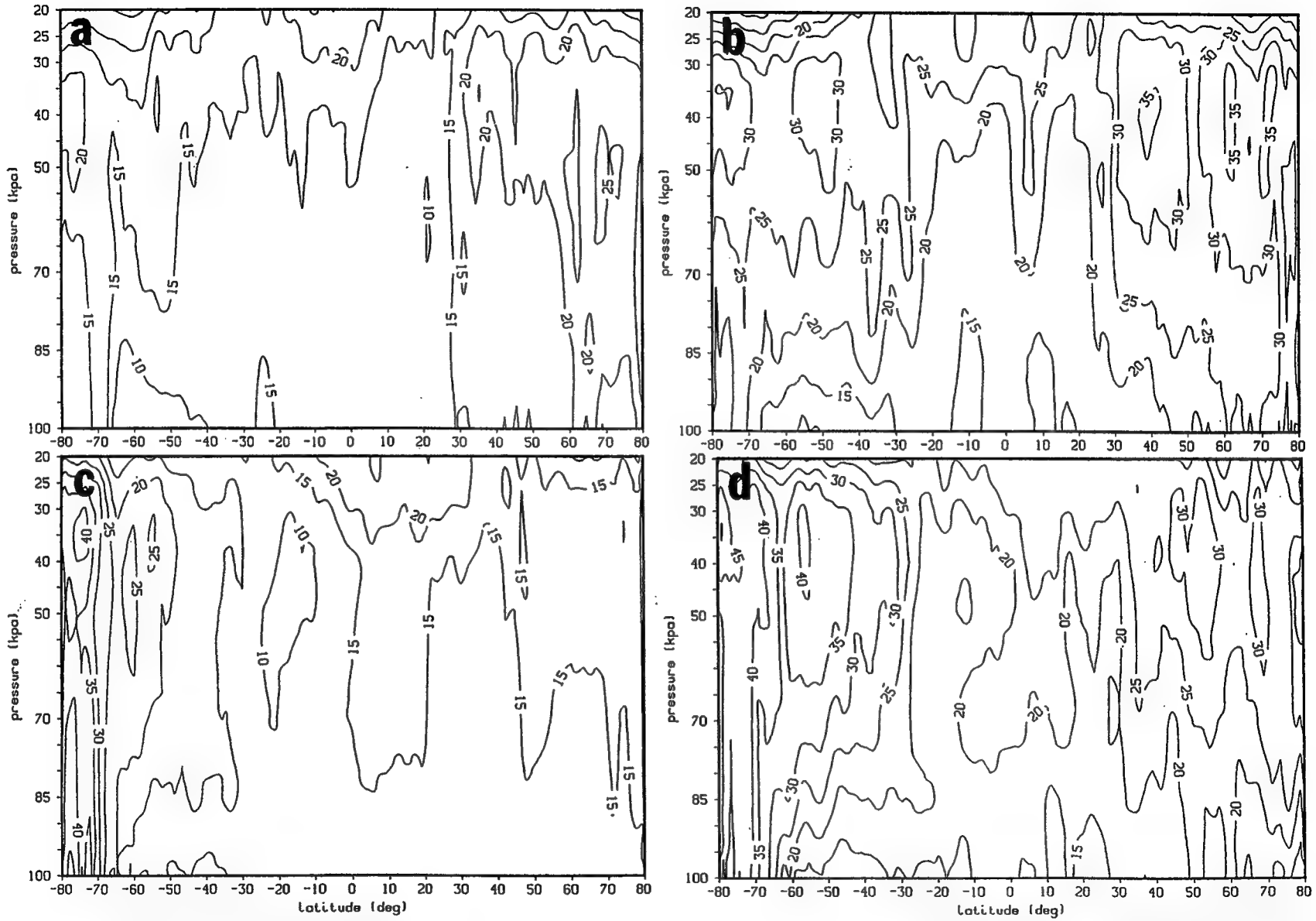

Figure 15. Zonal Cross Sections of RMSE of Relative Humidity (%) Computed on Indicated Pressure Surfaces at 1.125 Degree Latitude Intervals for T106L22 PL GSM (a) 12-Hour and (b) 48-Hour Forecasts Initialized 0000 UTC 15 January 1991, and (c) 12-Hour and (d) 48-Hour Forecasts Initialized 0000 UTC 15 July 1991.

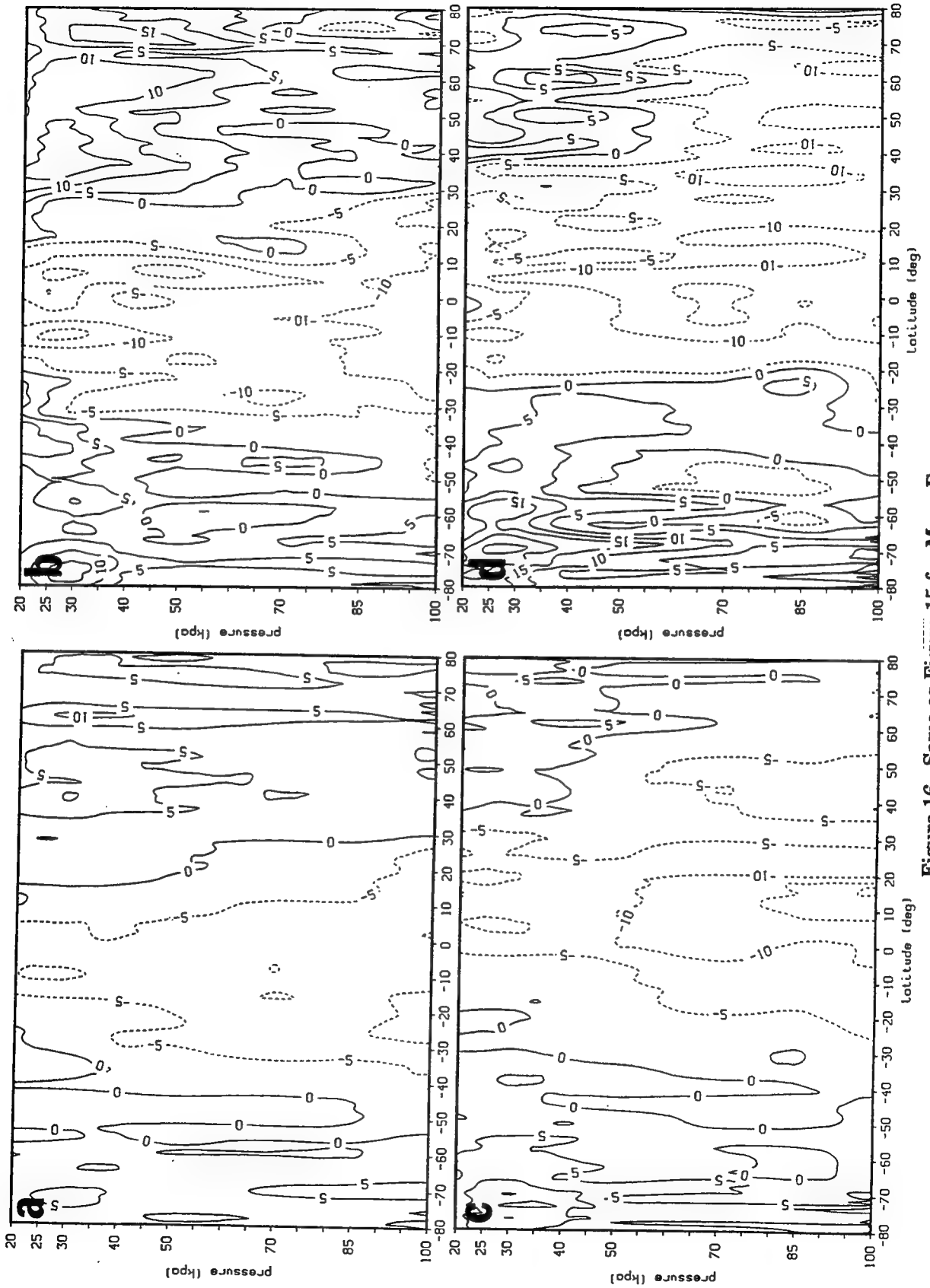

Figure 16. Same as Figure 15 for Mean Error.

CCA curves) would be available in the application of multiple linear regression (MLR) on the CCA-RTNEPH cloud amount differences. In addition, CCA cloud amounts were used as an additional predictor in the application of MLR cloud amounts. In the balance of this report, we shall refer to the application of MLR to the RTNEPH cloud amounts as MLR(RTN), and the application of MLR to the CCA-RTNEPH cloud amount differences as MLR(CCA-RTN).

3.5 Preparatory Experimentation

The CCA method relates frequency of occurrence of RH categories with frequency of occurrence of cloud amount to arrive at cloud amount--RH relationships. These relationships are often in tabular form, giving a single cloud amount value for each of a range RH values, from RH_c to RH_m. These tabular values are determined separately for each cloud deck, and may be plotted as cloud amount as a function of RH. The resulting curves are referred to as CCA curves by Trapnell[17], and we shall use that convention here as well.

To derive these relationships, we needed to determine which measure of forecast RH in a deck consisting of several NWP model σ layers might yield the best indicator of deck cloudiness. We tested both the RH_{ave} and RH_{max} as candidates for the CCA method. We first developed CCA curves based on a moving 7-day sample for application to period 8-24 January 1991 using both RH_{ave} and RH_{max}. Figure 17 compares the two sets of curves developed over the period 11-17 January. Not surprisingly, it is clear that for a given RH value, lesser cloud amounts are associated with curves based on RH_{max} than with curves based on RH_{ave}. This is because we matched the same cloud cover distributions with higher humidities in using RH_{max}.

We found that a majority of the gridpoints in the high cloud deck had RH_{max} = 100 percent. This is due at least in part to an upper tropospheric moist bias in the PL GSM, as documented by Norquist and Chang[34]. This led to many cases where less than 100 percent cloud cover points were associated with 100 percent RH_{max}, which explains the big jump in the high cloud RH_{max} curve at 100 percent RH. A similar

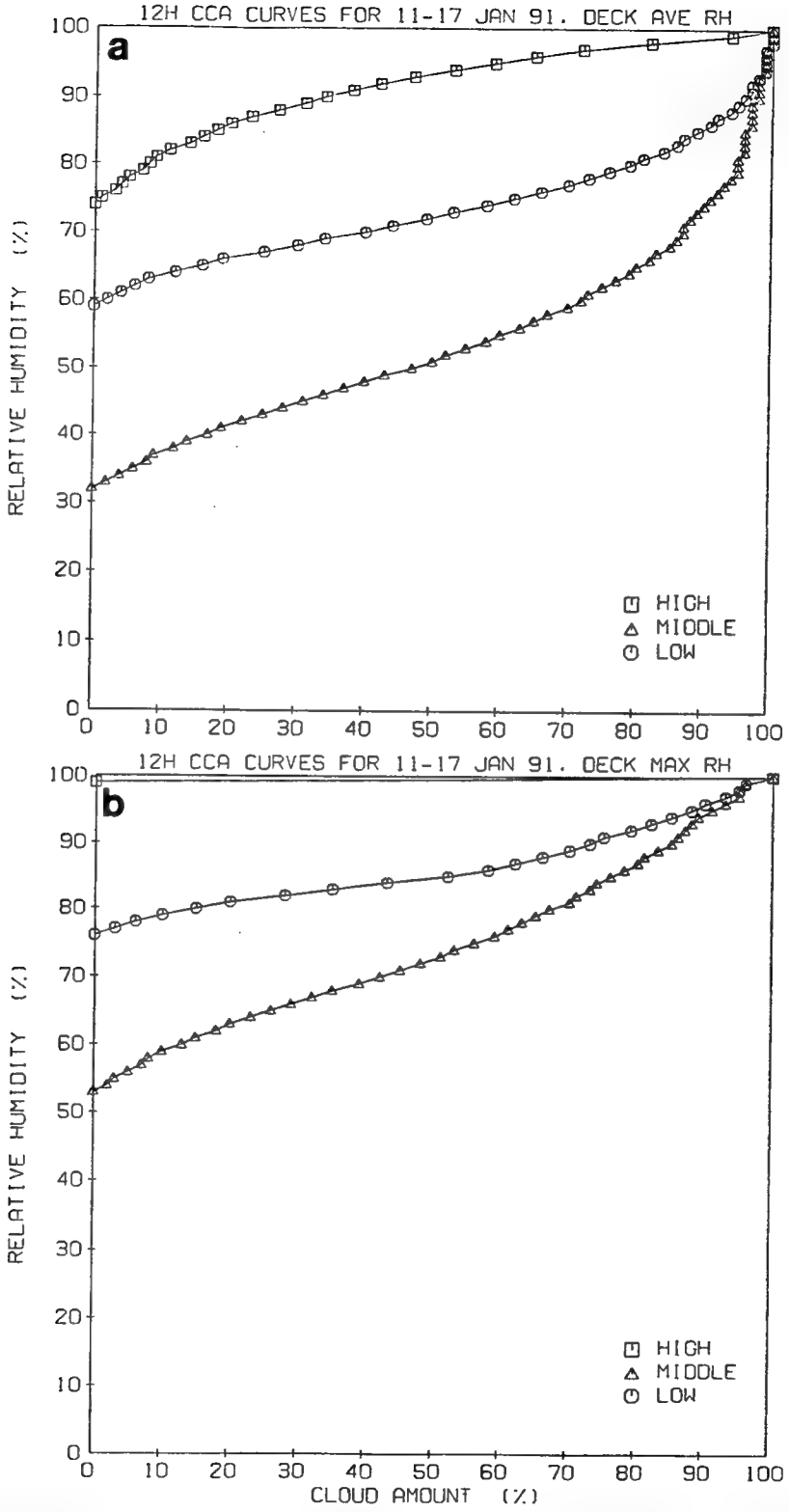

Figure 17. CCA Curves Developed From 12-Hour PL GSM Forecasts From Initial Times of 0000 and 1200 UTC 11-17 January 1991 of (a) Deck Averaged RH and (b) Deck Maximum RH and Valid Time Transformed RTNEPH Cloud Amounts.

feature was present in the July high cloud CCA curve using RH_{max}. The RH_{ave} distributions revealed many fewer gridpoints where high cloud deck RH = 100 percent, so no such jump in the cloud amounts is evident in the high cloud curve associated with RH_{ave}.

To make our final determination on the choice of RH to represent the NWP model forecasts in the CCA method, we simply verified a 7-day period of CCA forecasts in which both RH_{ave} and RH_{max} were used. The initial times for this set of 12-hour forecasts were 0000 UTC 18 January--1200 UTC 24 January 1991. We obtained larger biases, RMSEs, and mean absolute errors in all three decks using RH_{max}. This was particularly true for the high cloud, where the RMSE using RH_{max} was almost twice the value of RMSE obtained when RH_{ave} was used in the CCA method. We suspect that the latter result was due to the abundance of RH_{max} = 100 percent occurrences, leading to an inability of the scheme to distinguish various cloud high deck amounts associated with RH_{max} = 100 percent. We thus chose to practice the CCA cloud diagnostic method using cloud deck average RH to represent the NWP model forecasts.

After generating the representation of the predictand (RTNEPH cloud amount and CCA-RTNEPH cloud amount differences) and the potential predictors (see Table 1) on the transform grid, we began the process of developing predictor-predictand relationships. We started by applying the MLR procedure separately for each cloud deck over the entire hemisphere to the 12-hour PL GSM forecasts initialized twice daily on 8-17 January 1991 and RTNEPH cloud amounts (and separately, CCA-RTNEPH differences) at the 12-hour verifying times. We began the process using a screening regression algorithm that finds the reduction in variance for all 99 predictors, then starts eliminating predictors that have little or no contribution to the linear combination. However, we quickly found that this backward elimination procedure, applied for the 8-17 January data sample, swamped the relatively powerful mainframe computer we were using. We found that the greatest number of predictors that could be handled at one time by the system was about 40. To select this set of 40 predictors, we performed a forward selection screening procedure, in

which the single predictor (over the set of 99) that reduces the variance of the predictand more than any other predictor is identified first. Then, it chooses the predictor that, together with the first one selected, reduces the variance more than any other combination of two predictors. This process was continued until the top 40 predictors were identified. For the application of the MLR on the RTNEPH, Table 5 lists the top five predictors identified for the 8-17 January 12-hour forecasts. In the table, the figure 3X3 (3X3X3) represents the two-dimensional (three-dimensional) array of equal-area gridpoints centered on the subject gridpoint. Also included is the total correlation between the linear combination of the leading 20 predictors and the predictand. Note the prominence of RH (and RH^2), static stability ($\partial\theta/\partial Z$), and maximum boundary wind speed among the leading predictors. We found that the rate of reduction in variance with additional predictors becomes nearly zero after about the first 12 predictors.

Table 5. Five Leading Predictors [MLR(RTN)] in Each Deck
(Based on 8-17 January 1991 12-Hour Forecasts)

High Cloud

(Relative Humidity)^2 - High	Relative Humidity - Middle
Sine of Longitude	Precipitable Water - High
Wind Speed - Low	Total Cor (20 Pred) = 0.37

Middle Cloud

Maximum BL Wind Speed in 3X3	Relative Humidity - Middle
% Surface Covered by Water	Relative Humidity - High
Minimum Stability in 3X3X3	Total Cor (20 Pred) = 0.58

Low Cloud

Stability (t - 6) - Low	(Relative Humidity)^2 - Low
Minimum Stability in 3X3X3	Maximum BL Wind Speed in 3X3
Relative Humidity - Middle	Total Cor (20 Pred) = 0.55

The regression relationships obtained using the leading 20 predictors were

applied to the 12-hour PL GSM forecasts initialized at 0000 and 1200 UTC 18 January 1991. Figure 18 shows a plot of the resulting middle deck cloud amount forecasts valid 0000 UTC 19 January 1991 from MLR(RTN) and MLR(CCA-RTN). Also plotted for reference are the transformed RTNEPH and the 12-hour CCA forecast. Immediately obvious is the sharper cloud amount gradient in the CCA forecast, evidenced by well-defined extremes. In contrast, both MLR forecasts resulted in cloud distributions that are much flatter and poorly defined than the RTNEPH depiction. The MLR procedure minimizes the mean- squared error of estimate, at the expense of not preserving the true variance of the predictand's distribution. Thus, the greater extremes of cloudiness in the CCA forecast led to greater RMSEs than in the MLR forecasts, but produced sharper gradients.

We next conducted a series of experiments in which we modified the MLR process to increase the sharpness (percentage of points that are clear or nearly clear, or overcast or nearly overcast) in the MLR forecasts without significantly increasing the RMSE. In the case of MLR applied to RTNEPH, we modified the regression slope by factors of 2 and 3 in separate applications, to see if more cases of clear and overcast may be predicted. In the case of MLR applied to CCA-RTNEPH, we tried doubling the regression slope and excluding all humidity predictors in separate experiments. The philosophy behind the latter experiment was that, since the CCA cloud amount estimate is based totally on humidity, humidity is incorporated in the predictand. Thus, humidity would be too highly correlated with the predictand, possibly leading to the exclusion of other helpful predictors.

Table 6 gives the verification statistics for the MLR modification experiments. In each case, the predictor-predictand relationships were developed using the PL GSM 12-hour forecasts of the predictors initialized twice daily in the ten-day period 8-17 January, and were applied to the 12-hour PL GSM forecasts initialized 0000 and 1200 UTC 18 January 1991. In all of our objective verifications, we rounded forecast cloud amounts to the nearest 5 percent before verification. The quantity referred to as normalized sharpness (NS) in the table is the ratio of the sharpness of the forecast field to the sharpness of the observed field. We define sharpness in the same way as

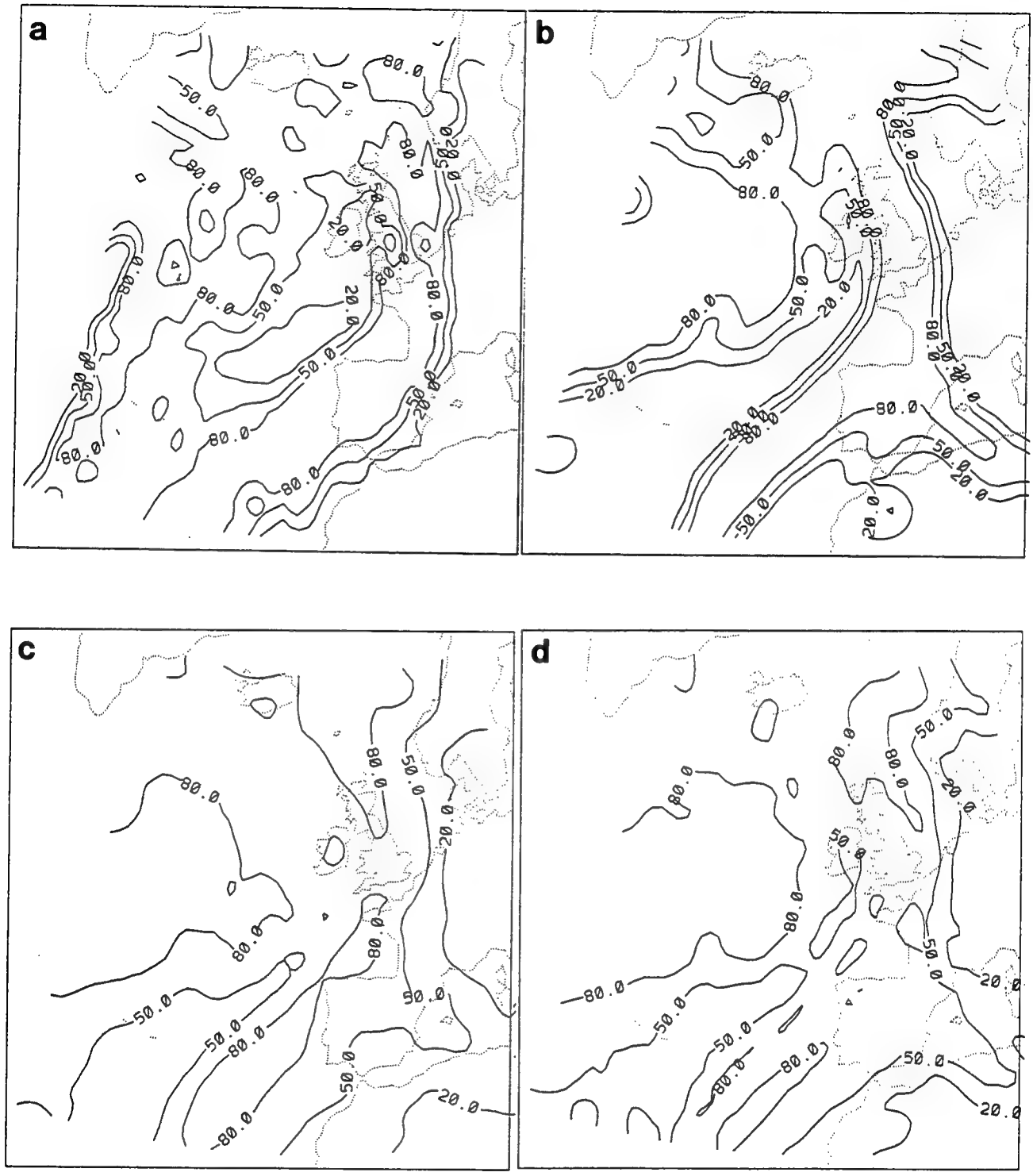

Figure 18. Middle Deck Cloud Amounts (%) for 0000 UTC 19 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast Using Standard MLR, (d) 12-Hour MLR(CCA-RTN) Forecast Using Standard MLR.

Cianciolo[22]--the fraction of the gridpoints having either no more than 20 percent or no less than 80 percent cloud amount over the verification sample. Ideally, we would want our forecasts to preserve the sharpness of the observed cloud fields--thus, a value of normalized sharpness closest to unity is most desirable. The other verification measure used in Table 6 is mean absolute error (MAE).

Table 6. NH Verification Statistics for Cloud Amount Forecasts
12-Hour Forecasts Initialized 0000, 1200 UTC 18 January 1991

		Cloud Deck					
Method	Slope Factor	Low		Middle		High	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
MLR (RTN)	1.0	0.1	31.0	0.7	31.7	0.2	15.4
MLR (RTN)	2.0	1.4	33.6	0.4	34.4	1.9	15.8
MLR (RTN)	3.0	2.6	36.4	0.5	37.0	4.3	17.8
MLR (CCA-RTN)	1.0	1.1	31.9	2.0	31.8	0.1	15.7
MLR (CCA-RTN)	2.0	4.4	39.5	7.6	41.9	2.5	18.2
MLR (CCA-RTN) w/o Hum.	1.0	2.3	38.6	1.5	36.8	0.7	19.1
		MAE	NS	MAE	NS	MAE	NS
MLR (RTN)	1.0	25.8	0.33	26.0	0.41	8.0	1.06
MLR (RTN)	2.0	25.3	0.75	24.9	0.84	8.6	0.97
MLR (RTN)	3.0	26.6	0.98	25.6	1.01	9.9	0.90
MLR (CCA-RTN)	1.0	26.4	0.35	26.2	0.39	8.1	1.06
MLR (CCA-RTN)	2.0	31.3	0.63	32.8	0.72	10.5	0.97
MLR (CCA-RTN) w/o Hum.	1.0	29.9	0.77	29.7	0.62	9.0	1.02

As stated previously, we sought to improve on the basic MLR procedure in such a way as to increase sharpness while minimally affecting forecast accuracy. We see that for both RTNEPH and CCA-RTNEPH predictands, increasing the regression slope increases sharpness and increases bias and RMSE. Interestingly, for MLR(RTN) applied to low and middle cloud, doubling the regression slope actually decreases the MAE while significantly increasing sharpness. For MLR(CCA-RTN), removing humidity variables as predictors increased the MAE less than doubling the regression slope, while still increasing the sharpness.

We feel that MAE may be a better indicator of forecast accuracy than RMSE for

this application, because in the latter extraordinary weight is given to larger misses (because of the squared factor). In applications where the range of the predictand is not limited, RMSE will detect extremely erroneous forecasts readily to indicate a significant problem in the forecast procedure or verification data. However, in our case, the predictand is rigidly limited to a fairly narrow range, making RMSE not as useful in detecting seriously erroneous forecasts. Furthermore, application of these forecasts for military decisions is made on a case by case, point by point basis. It is, therefore, important to minimize the typical, or average error rather than the root-mean-square error (which is better to minimize if forecast skill over ensembles of forecasts at many locations simultaneously must be optimized). For this reason, we chose the best balance between MAE and sharpness as our criterion for selecting the version of the MLR procedure to use in our experiments. On this basis, MLR(RTN) with a slope factor of 2, and MLR(CCA-RTN) without humidity predictors were chosen as the experimental MLR techniques for the predictands RTNEPH and CCA-RTNEPH respectively.

In addition to assessing the impacts of the modifications to the MLR procedure using verification statistics, it is instructive to examine the effects on the cloud amount categorical distribution. This can be done with a contingency table, which lists the percent frequency of occurrence of 5 percent category cloud amounts in the forecast fields vs. the observed fields. Figure 19 shows the contingency tables for the single-day verification period assessed above, for both the original and doubled regression slope application of MLR(RTN) for the middle cloud deck. The sum of the forecast percentages within 20 percent of the diagonal (the diagonal indicates cases of correct forecasts) are larger for the original slope table than for the doubled slope table in the 20-75 percent range of observed cloud amounts. In the extremes of the range of observed cloud, the double regression slope table indicates many more occurrences of cloud amount forecasts within 20 percent than did the original slope method. These results suggest that the doubled slope method achieves better sharpness than the original by better replicating the bimodality of the observed cloud cover distribution. This improvement comes at the expense of a slightly poorer skill

Fcst % **a**

100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	3
95	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2	1	4
90	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	2	1	5
85	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	2	2	5
80	1	0	0	0	0	0	0	1	0	0	0	1	0	1	1	1	1	2	3	2	6
75	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	3	3	2	8
70	1	1	1	0	0	1	1	1	1	1	1	1	1	2	1	2	2	3	3	3	8
65	3	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	9
60	5	1	1	1	1	1	1	1	1	1	1	2	1	2	1	2	2	3	3	2	7
55	8	1	2	1	1	1	1	1	2	1	1	2	2	2	2	3	2	3	5	4	7
50	11	2	2	2	3	2	2	2	2	2	3	2	2	2	2	3	2	3	4	3	7
45	16	3	3	2	3	2	2	3	2	2	2	2	2	3	2	2	3	3	3	2	7
40	23	4	5	3	4	2	3	3	3	2	3	2	2	3	2	3	3	3	3	2	6
35	33	6	5	4	4	3	3	3	2	3	2	3	2	2	2	3	3	3	3	2	7
30	41	8	6	4	4	3	3	4	3	3	2	2	2	2	2	2	2	2	2	2	4
25	41	8	5	4	4	3	3	3	2	2	2	2	1	2	1	2	1	1	1	1	4
20	38	6	4	3	3	2	2	2	1	2	1	1	1	1	1	1	1	1	1	1	3
15	28	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	2
10	22	3	3	2	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1
5	18	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	27	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100

Fcst % **b**

100	2	1	1	1	1	1	1	2	1	3	2	5	3	5	4	7	7	9	15	9	36
95	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	4
90	1	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	2	1	5
85	2	0	1	1	0	0	0	1	1	0	1	1	0	1	1	1	1	1	2	1	5
80	2	1	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	2	1	4
75	3	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2	1	2
70	4	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	2	3	2	4
65	5	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3
60	5	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	4
55	6	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	3
50	7	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	3
45	10	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	3
40	11	2	2	2	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	3
35	13	2	3	1	2	2	1	2	1	1	1	1	1	2	1	2	2	1	2	1	3
30	16	3	3	2	2	1	2	1	1	2	1	1	1	1	1	2	2	1	2	1	4
25	18	4	3	2	2	2	1	1	1	1	1	1	1	1	1	2	1	1	2	1	3
20	21	4	3	2	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	0	2
15	21	4	3	2	2	2	2	2	1	1	1	1	1	1	1	1	0	1	1	1	2
10	21	4	3	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	2
5	18	3	2	2	2	1	1	2	1	1	1	1	1	1	1	1	0	1	1	1	1
0	127	14	12	7	6	4	3	4	3	4	3	4	3	2	2	2	2	2	2	1	6
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100

Obs %

Figure 19. Middle Cloud Deck Contingency Table (Percent X 10 Frequency of Occurrence of Cloud Amount Diagnoses at Gridpoints Falling with Simultaneous Forecast and Observed 5% Categories) for 12-Hour Forecasts Initialized 0000, 1200 UTC 18 January 1991 Using (a) MLR(RTN) with Standard (1.0) Regression Slope, (b) MLR(RTN) with Doubled (2.0) Regression Slope.

of the doubled slope method in intermediate cloud amount categories. However, as seen in Figure 8, these intermediate categories (20-75 percent) represent a minority (about 30 percent for January RTNEPH) of the observed cloud amount cases. Thus, it seems wise to select the method that gives better forecasts in cloud amount cover categories occurring more of the time (70 percent) to increase our probability for success. For this reason, we chose the doubled slope method over the unmodified slope method for MLR(RTN). For MLR(CCA-RTN), similar observations were made when comparing contingency tables (not shown) for the original method and the method in which humidity predictors were withheld. For this reason, the latter method was chosen for the CCA-RTNEPH predictand. From this point forward in the report, references to MLR(RTN) and MLR(CCA-RTN) will imply the use of these chosen modifications to the original procedure.

3.6 Experimental Design

The CCA cloud-forecast RH relationships were developed over moving 7-day periods in January and July 1991. The first development period in each month spanned the period 0000 UTC of day 1 to 1200 UTC of day 7. Deck average PL GSM RH forecasts initialized at each of these times were used with transform grid RTNEPH cloud amounts at the forecast verification times. Resulting cloud amount--RH relationships (the CCA curves) were applied to the PL GSM RH forecast initialized at 0000 UTC on day 8. That is, we applied each developed set of CCA curves to a single initial forecast time. Thus, the development period moved forward a half-day at a time, as does the application date and time. The last development period in each month spanned the period 1200 UTC day 17 to 0000 UTC day 24, and consequent relationships were applied to the 1200 UTC day 24 forecast. We felt that we could avoid a two-week development period, like that used by Trapnell[17], because the entire hemisphere is used (unlike Trapnell's regional CCA curves).

We developed the NH predictor-predictand relationships in each cloud deck for both the MLR(RTN) and MLR(CCA-RTN) over moving 10-day periods in January and

July 1991. The first ten-day development period in each month involved PL GSM forecasts initialized at 12-hour intervals between 0000 UTC of day 8 and 1200 UTC of day 17 inclusive, along with RTNEPH cloud amounts on the transform grid at the corresponding PL GSM forecast verification times. The resulting cloud amount-NWP model predictor relationships were then applied to PL GSM forecasts initialized at 0000 and 1200 UTC on day 18. This pattern of a 10-day development period followed by application on the very next day was conducted for 7 consecutive days (which we shall refer to as the verification period) ending with a day 14 to day 23 development period and a day 24 application. We based our choice of this approach on the hypothesis that the most representative sample of weather over which to design a regression relationship is the weather immediately prior to the forecast initialization date. The large forecast region (the hemisphere) allows us to use a rather limited temporal sample. This is because at any synoptic time, a number of weather and attendant cloud scenarios exist in the hemisphere. In the 10-day development ensemble, these scenarios should encompass any weather-cloud pattern seen in the hemisphere on the application day.

We verified the ensemble of cloud diagnoses from PL GSM forecasts initialized on days 18-24 using all three methods--CCA, MLR(RTN), MLR(CCA-RTN)--against transformed RTNEPH cloud amounts at forecast verification times. In addition to the three previously-mentioned verification measures (bias, RMSE, MAE), we followed Cianciolo[22] in computing Brier and 20/20 scores and sharpness. All three of these statistical quantities are derived from contingency tables, which we constructed for each deck and each forecast duration. In addition to hemispheric statistics, we conducted verifications and constructed contingency tables for the four separate subregions: 0-30N land and water surfaces, and 30-90N land and water surfaces. Once again, it should be stressed that while the verification was both hemispheric and regional, the cloud diagnosis development and application procedures were fully hemispheric only. As has been mentioned previously, it may be profitable in future work to investigate regional development and application of the cloud diagnosis procedures.

4. EXPERIMENTAL RESULTS

4.1 CCA Curves

The CCA cloud amount-RH relationships, or CCA curves, were developed and used to diagnose cloudiness at 12-hour intervals from 0000 UTC of day 8 until 1200 UTC of day 24 of both months. As mentioned earlier, preliminary experiments showed that the CCA method produced better diagnosis of cloudiness when the deck average RH was used than when deck maximum RH was used. Therefore, in all of the results involving CCA described below, we used the deck average RH to represent the PL GSM forecasts.

In Figure 20, we show the CCA curves relating the 12-hour PL GSM RH forecasts with the transformed RTNEPH cloud amounts at the verification times for 7-day development periods beginning 0000 UTC on 1 and 17 January 1991. Figure 21 shows the same curves for the corresponding July periods. We chose to display these periods because they represent the first and nearly last development periods in our data sample. We observe a significant difference between the curves for the two development periods in middle and low decks in January and in the high deck in July. Use of the 01-07 Jan middle and low deck curves for 17-23 Jan would lead to an underestimation of cloud amount by as much as 12 percent. Similarly, using the earlier July high cloud curve for the latter period would produce too little cloud by as much as 7 percent.

When we compare CCA curves of 12-hour forecasts with those of 48-hour forecasts in each month (Figures 22 and 23), we observe some noticeable differences. In January, the high and low deck curves shift toward lower humidities, while the middle deck moves to higher RH. This suggests that the PL GSM hemispheric RH is decreasing in the high and low decks and increasing in the middle deck between 12 and 48 hours of forecast time. In contrast, in July, the RH in the low deck is increasing with forecast time while the middle deck experiences little change. This appears to be consistent with the zonal cross sections of RH bias shown in Figure 16.

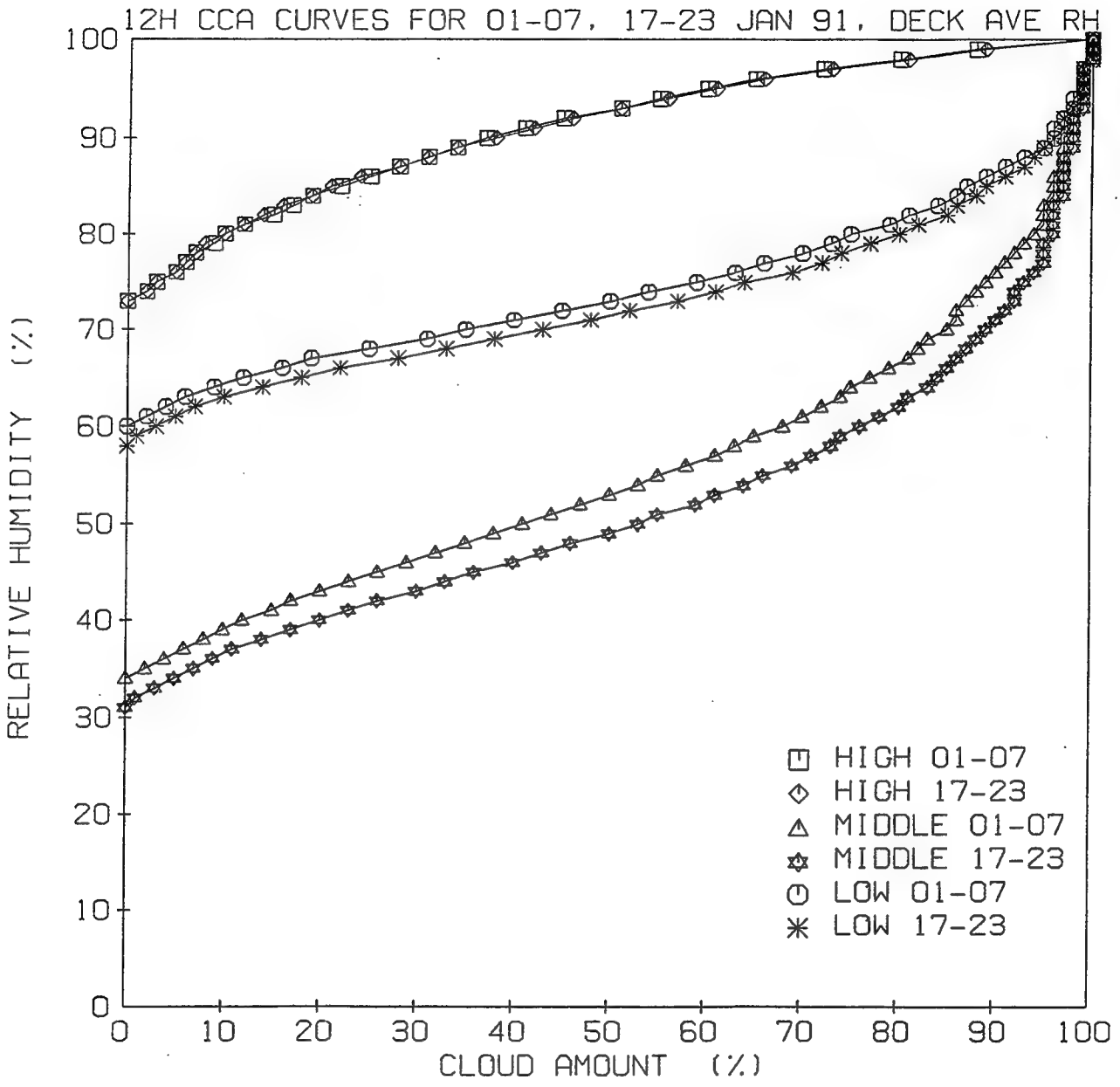

Figure 20. **CCA Curves Relating 12-Hour PL GSM RH Forecasts With the Transformed RTNEPH Cloud Amounts at the Verification Times for Seven-Day Development Periods Beginning 0000 UTC 1 and 17 January 1991.**

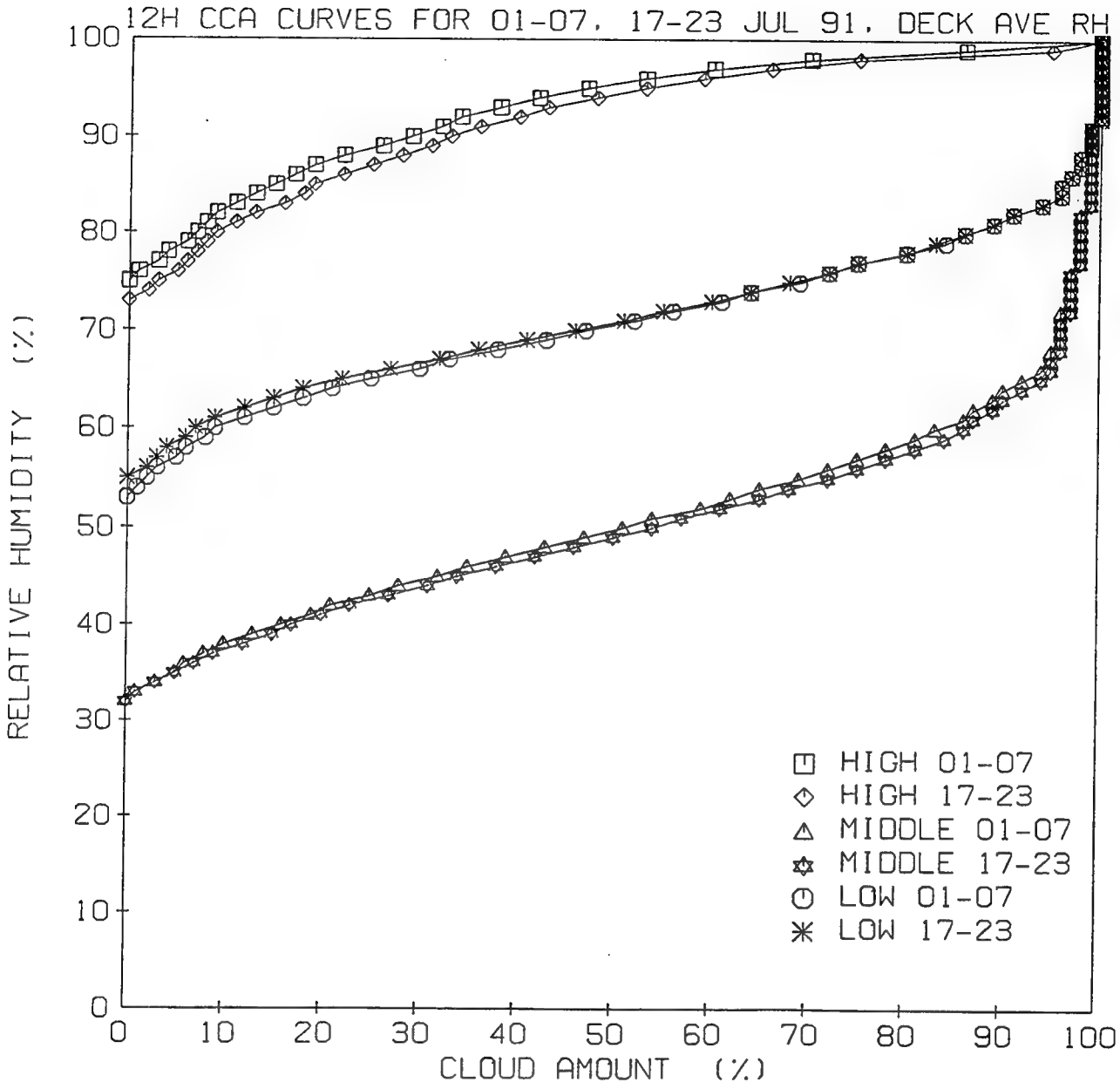

Figure 21. Same as Figure 20 for July Periods.

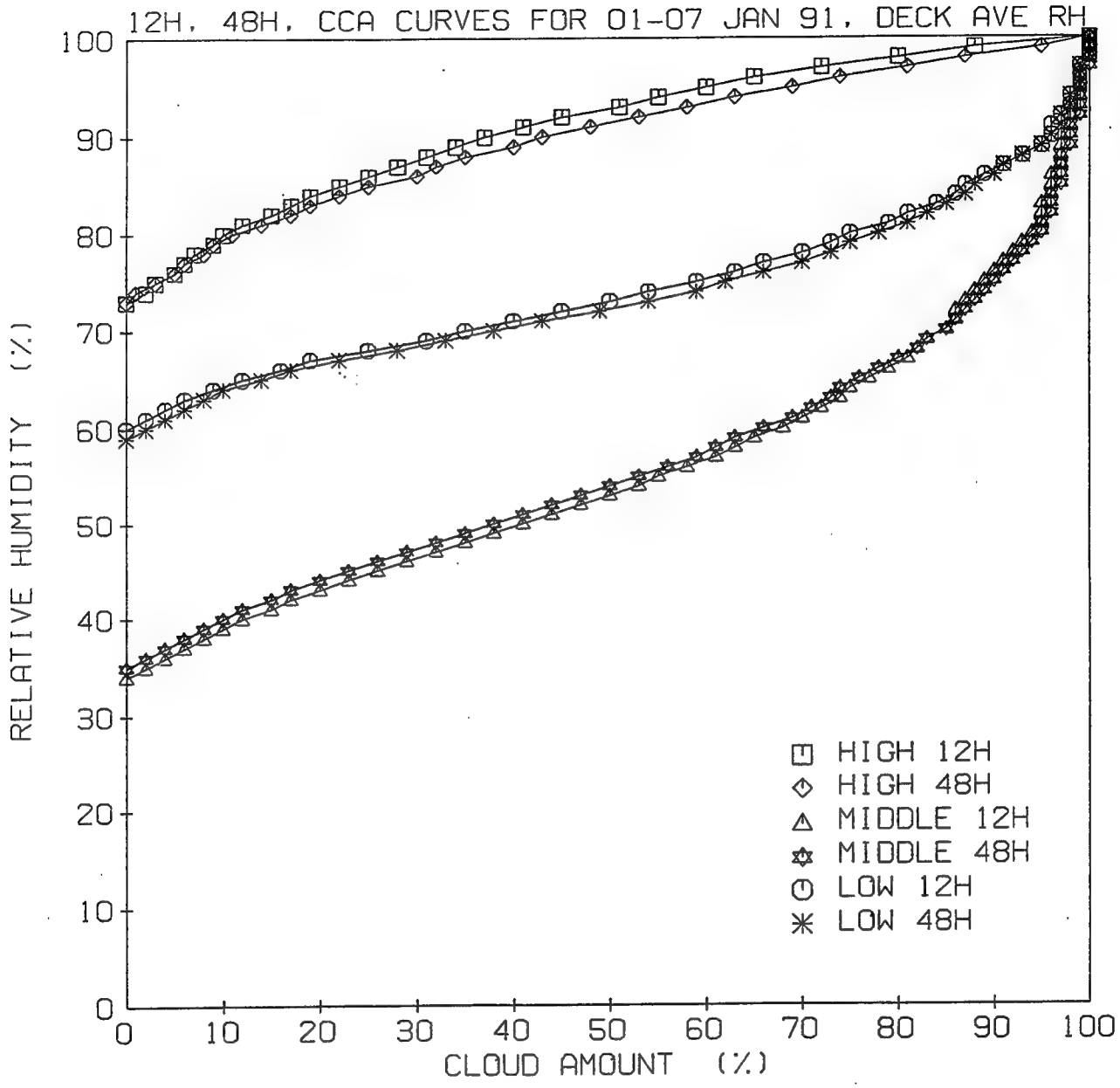

Figure 22. CCA Curves Relating 12-Hour and 48-Hour PL GSM RH Forecasts With the Transformed RTNEPH Cloud Amounts at the Verification Times for a Seven-Day Development Period Beginning 0000 UTC 1 January 1991.

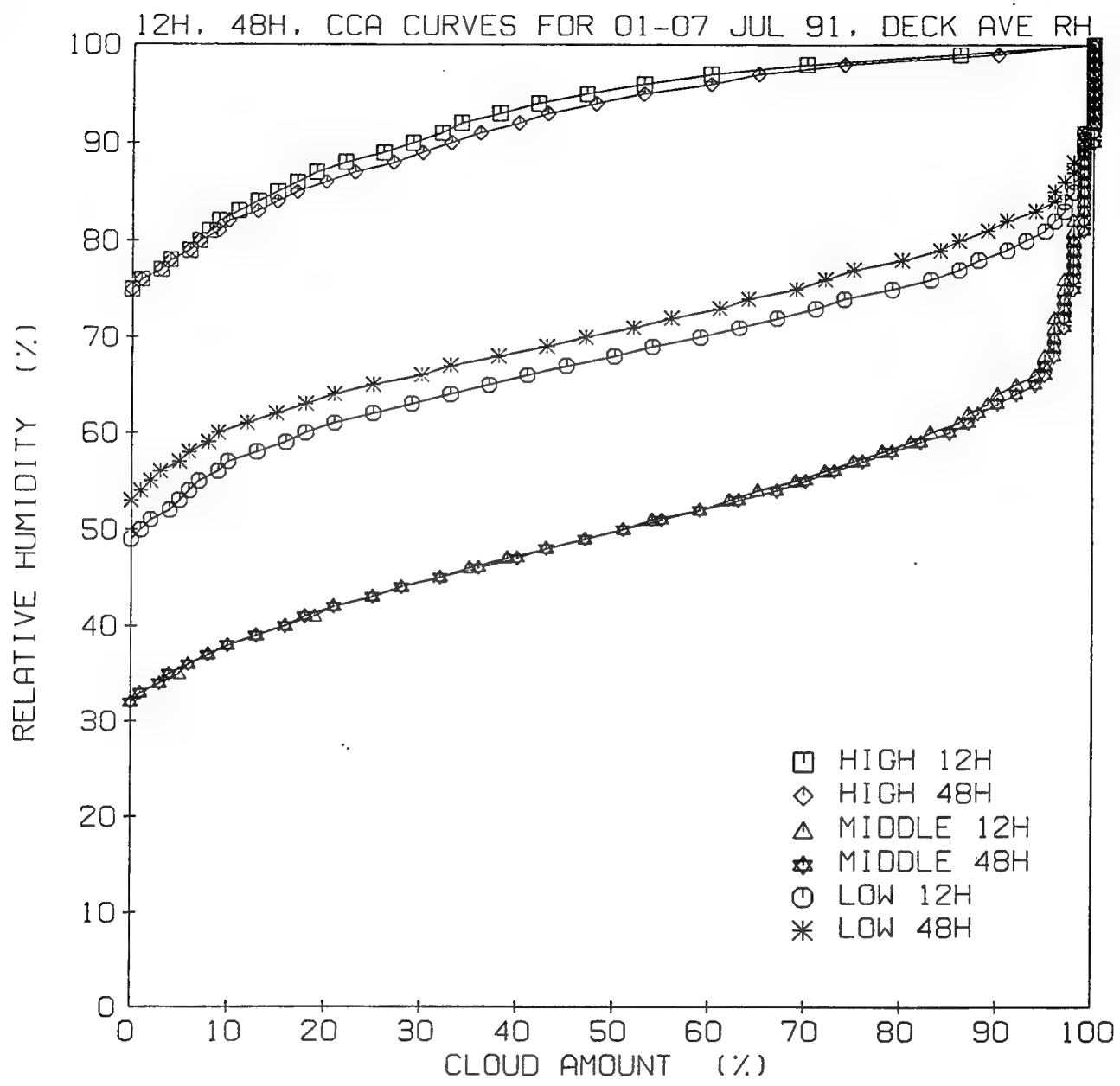

Figure 23. Same as Figure 22 for the July Period.

Note in particular the decrease of negative (dry) low-level tropical bias between 12 and 48 hours in July. Finally, the differences between January and July CCA curves reflect seasonal differences in both forecast RH and cloud amount distribution.

4.2 Cloud Amount Predictors for 12-Hour Forecasts

We now turn our attention to the variables identified as leading predictors by the MLR procedure. As previously stated, for each month, forecast duration, and cloud deck, we developed predictand-predictor relations for each of seven 10-day periods. These were the 10-day sets immediately prior to each of the seven days in the verification period, days 18-24 of each month. For each of the seven development periods, we looked at how the MLR(RTN) scheme prioritized the leading predictors for that sample, selected on the basis of correlation (for the first selectee) and multiple correlation with cloud amount. For the purposes of discussion, we devised a simple scoring scheme in which a variable identified in the top 10 predictors in a given development period is assigned two points, and variables in the second 10 predictors are assigned one point. We added the point assignments over the seven development periods and designated two predictor categories: strong predictors (total points $\geq$10) and useful predictors (10> total points >5). Incidentally, for a given month, forecast duration, and cloud deck, we observed little day-to-day change in the identified predictors. We did see some minor changes in their order of priority.

In Table 7, we list the strong and useful predictors for the three decks according to four categories: dynamic, humidity, geographic, and turbulence. In the 12-hour forecasts of both months, we see that RH is a strong predictor in all three decks. Apart from RH, there are significant differences in the strong predictors between the two months. For example, dry static stability is a strong predictor in all three decks in July, but only in the low deck in January, when moist static stability is strong in low and high decks. Similarly, minimum stability in a 3X3X3 gridpoint array is a strong low and middle deck predictor in January, but doesn't appear at all in July's leading predictors. Geographic predictors play a much greater role in July 12-hour

Table 7. Strong (x) and Useful (+) Predictors in 12-Hour MLR (RTN)

Dynamic	L	M	H	Humidity	L	M	H	Geographic	L	M	H	Turbulence	L	M	H
JANUARY															
vorticity		x		relative humidity	x	x	x	sin (longitude)			x	dry static stab	x	+	+
temperature		+	x	(rel humid)^2	x		+	cos (longitude)	+	+	+	moist static stab	x	+	x
zonal wind	+			(rel humid)^4	x		+	hrs. sun before t		+	+	wind speed	+	+	x*
meridional wind	+	+	x	precip water	+	+		hrs dark before t	+			wind shear			+
surface pressure	+			ice rel humid	+	+	+	% surface water	x	x		ln (Ri)			+
vertical velocity			+	cond pres spread		+						3X3X3 min stab	x	x	
				lifting cond level	+							3X3X3 max speed		x	
												3X3 BL max speed	x	x	+
												evaporation rate	+	x	+
												3X3 max conv prcp	+	+	
												Z. var X BL max spd		+	
												* speed in low deck			
JULY															
temperature	x			relative humidity	x	x	x	sin (latitude)	x	+	x	dry static stab	x	x	x
meridional wind		x	x	(rel humid)^4	+			latitude			+	wind speed	+		
vertical velocity	+	+		precip water		x		sin (longitude)	+		x	wind shear	+	x	
moist convergence		+		cond pres spread	x			cos (longitude)		x		3X3X3 max speed		+	
stratiform prcp			+					cos (zenith angle)	x			3X3 BL max speed		+	
								hrs sun before t	+		+	evaporation rate	x	+	x
								hrs dark before t	+	x	x	3X3 max conv prcp			x
								% surface water			x				
								surface height	+	+					

forecasts. Evaporation rate is a useful predictor in every deck for both months, and is a strong predictor in three cases. Over all three decks, the five strongest predictors for 12-hour forecasts are: January--relative humidity, moist static stability, maximum boundary layer wind speed in a 3X3 array, percent surface water, and minimum dry static stability in a 3X3X3 array; July--relative humidity, dry static stability, evaporation rate, hours of darkness before forecast valid time t, and sine of latitude.

One of the stated purposes of this study was to determine the degree to which other variables may supplement a single humidity variable as predictors of cloud amount. To do this, we compared the cloud amount correlation of the most highly correlated humidity variable with the total correlation of the 20 leading predictors. The increase in correlation gives a quantitative measure of the amount of additional information on cloud amount distribution available in the supplemental variables. Among humidity variables, the single most highly correlated variable and their correlations for January are: low--RH^2 (0.35), middle--RH (0.32), high--RH^4 (0.26); for July: low--condensation pressure spread at t-6 hours (-0.32); middle--precipitable water (0.42), high--RH (0.17). We computed the 7-day average of the total correlation of the linear combination of the top 20 predictors. The average values for January were: low--0.55, middle--0.58, high--0.39; for July: low--0.44, middle--0.58, high--0.34. On a percentage basis, the supplemental predictors increased the correlation most in the July high deck, and least in the July low and middle decks.

4.3 12-Hour Cloud Forecast Verification Statistics

Next, we look at the statistical verification scores for the 12-hour cloud forecasts. These forecasts were diagnosed by applying each of the seven single-day predictand-predictor relationship sets to the corresponding verification period day's PL GSM forecast. Tables 8 and 9 list the scores computed for the ensemble of 14 twice-daily forecasts made over the 7-day verification period in January and July respectively. In the following discussion, we will first focus on just the verification results for the entire hemisphere (0-90N in Tables 8 and 9). We choose to emphasize the verification

over the whole hemisphere because, for the CCA, MLR(RTN), and MLR(CCA-RTN) procedures, this was the region of development. We can only be sure that the performance of each of the procedures is truly representative of the actual skill over the region of their development. However, their verification in subregions can give us information on the unique characteristics of the subregion and the potential for improvement in subregional applications of the procedures. We shall discuss the subregional results following our discussion of the hemispheric results.

Given the nature of the CCA and MLR procedures, we expect that, when verified over the region of their development, their forecast biases will be small. We see from the tables that bias exceeds 2 percent cloud amount only in July low clouds for CCA and MLR(RTN). Such small values suggest that there is no systematic error (repeated errors of the same sign) in the cloud forecast procedures.

As previously mentioned, we feel that mean absolute error (MAE) may be the single best measure of forecast skill for point-by-point forecast applications. MAE identifies the average departure of the point forecast from the observed cloud amount, regardless of sign. For example, a MAE of 20 percent cloud cover would indicate that for a forecast at any point, on the average the forecasted value would differ from the verifying observed value by 20 percent cloud amount.

The most striking feature of the MAE results is the much lower error in the high deck. We see that this is true for all score categories in the table beginning with MAE. The primary reason for this can be seen in Figure 8, the transformed RTNEPH cloud amount distributions. Notice that there are many more cases of no cloud in the high deck, and the remaining cloudy cases are largely restricted to lower cloud amounts. Thus, high cloud is much more nearly a binary predictand than are low and middle clouds (with "no" being much more likely than "yes") and cloudy high deck cases are restricted to a more limited range of cloud amounts. All forecast methods perform at their best under such conditions, because any forecast of clear or nearly clear is very likely to be correct. We believe that the much higher skill associated with high clouds in our study is due at least partly to the probable underspecification of high cloud amounts in our study, as discussed in Section 3.3.

Table 8. 12-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jan 91 -1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			Bias (% Cloud Amount)			
High	CCA	0.6	0.8	-5.1	0.8	-0.8
	MLR(RTN)	2.7	0.6	0.6	3.5	1.6
	MLR(CCA-RTN)	2.8	1.3	-1.3	0.9	0.6
	Persistence	0.0	-0.1	-0.2	-0.2	-0.1
	Random	46.9	47.8	40.1	41.7	43.9
Middle	CCA	2.7	-15.5	26.4	-7.5	-0.2
	MLR(RTN)	-9.1	2.5	-8.8	5.4	-0.9
	MLR(CCA-RTN)	1.0	2.8	5.1	-3.6	1.5
	Persistence	0.8	-0.6	0.5	-0.6	-0.1
	Random	23.8	9.0	23.6	-9.1	9.3
Low	CCA	-3.7	-7.9	16.1	0.5	1.4
	MLR(RTN)	-9.2	4.2	-8.6	8.0	0.5
	MLR(CCA-RTN)	-1.6	1.8	4.8	-0.5	1.6
	Persistence	0.6	-0.7	0.6	-0.1	0.0
	Random	22.6	6.0	20.5	-5.1	8.4
			Mean Absolute Error (% Cloud Amount)			
High	CCA	6.1	4.4	12.3	10.8	8.5
	MLR(RTN)	7.3	4.0	13.2	12.6	9.1
	MLR(CCA-RTN)	7.7	4.9	13.4	11.5	9.3
	Persistence	4.2	3.8	13.4	11.9	9.0
	Random	48.3	49.1	45.2	47.1	47.4
Middle	CCA	22.6	29.1	38.6	37.3	33.3
	MLR(RTN)	20.2	26.0	23.1	27.0	24.9
	MLR(CCA-RTN)	23.1	29.2	27.4	30.3	28.4
	Persistence	21.0	25.8	25.0	25.2	24.9
	Random	43.2	41.1	41.9	39.1	41.0
Low	CCA	23.6	31.6	33.6	33.1	31.7
	MLR(RTN)	22.0	28.2	23.1	24.5	25.2
	MLR(CCA-RTN)	23.4	30.4	27.6	27.7	28.2
	Persistence	24.9	31.3	24.3	25.0	26.8
	Random	40.1	37.3	41.4	38.6	39.0

Table 8. 12-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jan 91 - 1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			RMSE (% Cloud Amount)			
High	CCA	17.4	15.2	25.2	24.1	21.1
	MLR(RTN)	13.1	11.4	20.5	19.2	16.7
	MLR(CCA-RTN)	15.5	13.0	23.0	22.5	19.2
	Persistence	12.7	14.9	25.1	26.9	22.0
	Random	56.3	57.1	53.5	55.2	55.5
Middle	CCA	34.0	40.8	50.1	48.3	45.1
	MLR(RTN)	31.6	34.4	33.8	36.3	34.5
	MLR(CCA-RTN)	32.2	35.7	36.3	39.2	36.6
	Persistence	33.4	37.4	37.8	35.0	36.4
	Random	51.7	49.5	50.3	47.3	49.4
Low	CCA	34.6	39.9	46.5	44.0	42.5
	MLR(RTN)	31.3	34.3	33.8	33.6	33.7
	MLR(CCA-RTN)	32.3	36.7	37.7	36.7	36.6
	Persistence	35.3	40.7	36.8	34.7	37.3
	Random	48.4	45.5	49.8	46.9	47.4
			Brier Score (perfect = 0)			
High	CCA	0.031	0.023	0.064	0.058	0.045
	MLR(RTN)	0.017	0.013	0.042	0.037	0.028
	MLR(CCA-RTN)	0.024	0.017	0.053	0.051	0.037
	Persistence	0.016	0.022	0.063	0.073	0.048
	Random	0.317	0.326	0.286	0.305	0.308
Middle	CCA	0.116	0.166	0.251	0.233	0.203
	MLR(RTN)	0.100	0.118	0.114	0.132	0.119
	MLR(CCA-RTN)	0.104	0.127	0.132	0.154	0.134
	Persistence	0.112	0.140	0.143	0.123	0.133
	Random	0.267	0.245	0.253	0.224	0.244
Low	CCA	0.119	0.159	0.216	0.194	0.181
	MLR(RTN)	0.098	0.117	0.114	0.113	0.113
	MLR(CCA-RTN)	0.104	0.134	0.142	0.135	0.134
	Persistence	0.124	0.166	0.135	0.121	0.139
	Random	0.235	0.207	0.248	0.220	0.224

Table 8. 12-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jan 91 - 1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			20/20 Score (perfect = 1)			
High	CCA	0.901	0.934	0.800	0.825	0.864
	MLR(RTN)	0.929	0.962	0.815	0.822	0.880
	MLR(CCA-RTN)	0.915	0.950	0.827	0.844	0.883
	Persistence	0.930	0.944	0.776	0.821	0.858
	Random	0.242	0.236	0.275	0.257	0.253
Middle	CCA	0.618	0.528	0.413	0.418	0.475
	MLR(RTN)	0.651	0.529	0.598	0.525	0.558
	MLR(CCA-RTN)	0.599	0.437	0.489	0.469	0.476
	Persistence	0.655	0.577	0.602	0.567	0.589
	Random	0.298	0.320	0.312	0.339	0.321
Low	CCA	0.598	0.437	0.485	0.474	0.476
	MLR(RTN)	0.606	0.442	0.594	0.561	0.532
	MLR(CCA-RTN)	0.581	0.414	0.515	0.519	0.487
	Persistence	0.563	0.462	0.608	0.567	0.545
	Random	0.326	0.359	0.315	0.345	0.340
			Normalized Sharpness (perfect = 1)			
High	CCA	1.005	0.985	1.100	0.982	1.015
	MLR(RTN)	1.016	1.006	1.009	0.892	0.976
	MLR(CCA-RTN)	1.007	0.997	1.105	0.994	1.024
	Persistence	0.999	1.001	1.006	1.005	1.003
	Random	0.475	0.463	0.525	0.492	0.488
Middle	CCA	0.899	1.004	0.928	1.079	0.991
	MLR(RTN)	0.926	0.714	0.889	0.931	0.842
	MLR(CCA-RTN)	0.774	0.480	0.609	0.776	0.623
	Persistence	0.994	1.006	0.997	1.001	1.001
	Random	0.569	0.622	0.593	0.680	0.623
Low	CCA	1.130	0.784	1.050	1.098	0.986
	MLR(RTN)	0.966	0.380	0.846	1.058	0.766
	MLR(CCA-RTN)	0.912	0.434	0.758	0.885	0.704
	Persistence	0.995	1.002	0.999	1.000	1.000
	Random	0.835	0.830	0.883	0.770	0.826

Table 9. 12-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jul 91 - 1200 UTC 24 Jul 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			Bias (% Cloud Amount)			
High	CCA	1.2	-0.7	-2.8	3.4	0.2
	MLR(RTN)	5.5	1.3	2.7	-0.2	1.6
	MLR(CCA-RTN)	2.4	1.0	0.3	2.6	1.4
	Persistence	0.2	0.4	-0.2	0.0	0.1
	Random	43.3	45.0	40.8	47.5	44.5
Middle	CCA	-0.7	-16.8	15.8	6.5	0.0
	MLR(RTN)	3.3	6.5	-3.1	-0.8	1.6
	MLR(CCA-RTN)	2.7	0.3	4.1	1.6	1.9
	Persistence	-0.3	-0.7	-0.8	0.6	-0.2
	Random	-3.5	6.2	14.3	6.0	3.0
Low	CCA	-4.6	-2.4	-9.4	22.7	3.3
	MLR(RTN)	3.3	7.7	-3.7	0.5	2.2
	MLR(CCA-RTN)	0.3	0.9	0.4	4.6	1.8
	Persistence	-0.2	0.2	-0.9	0.3	0.0
	Random	8.6	-1.5	8.9	7.8	4.9
			Mean Absolute Error (% Cloud Amount)			
High	CCA	12.0	7.4	12.2	7.4	9.1
	MLR(RTN)	12.3	8.3	11.7	4.0	8.2
	MLR(CCA-RTN)	11.8	8.9	12.7	6.7	9.5
	Persistence	10.1	7.8	11.6	4.1	7.6
	Random	46.6	48.3	44.7	48.6	47.3
Middle	CCA	29.2	33.5	32.6	32.1	32.5
	MLR(RTN)	24.1	26.2	22.6	25.7	25.0
	MLR(CCA-RTN)	27.8	29.2	26.9	28.8	28.4
	Persistence	26.0	25.2	26.7	29.3	27.1
	Random	42.7	42.1	40.5	40.2	41.1
Low	CCA	31.1	34.6	30.0	37.8	34.1
	MLR(RTN)	26.0	30.2	23.5	27.4	27.3
	MLR(CCA-RTN)	27.3	31.3	27.6	29.8	29.6
	Persistence	31.0	36.8	30.1	30.9	32.6
	Random	37.3	37.7	38.2	37.0	37.6

Table 9. 12-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jul 91 - 1200 UTC 24 Jul 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			RMSE (% Cloud Amount)			
High	CCA	25.9	21.4	23.9	18.8	21.8
	MLR(RTN)	18.2	17.0	16.8	9.9	15.3
	MLR(CCA-RTN)	20.8	19.0	21.7	15.9	19.0
	Persistence	22.6	22.0	21.6	12.6	19.2
	Random	54.8	56.3	53.0	56.6	55.4
Middle	CCA	44.1	45.3	44.4	43.7	44.3
	MLR(RTN)	34.6	36.8	31.5	34.8	34.7
	MLR(CCA-RTN)	36.5	37.7	36.0	38.2	37.3
	Persistence	39.0	37.8	37.9	40.5	38.9
	Random	50.9	50.4	48.8	48.5	49.5
Low	CCA	41.4	42.4	41.2	47.4	43.6
	MLR(RTN)	34.6	37.3	31.2	34.1	34.6
	MLR(CCA-RTN)	35.4	37.7	36.2	36.6	36.8
	Persistence	41.3	46.8	40.2	40.1	42.5
	Random	45.4	45.9	46.5	45.1	45.8
			Brier Score (perfect = 0)			
High	CCA	0.067	0.046	0.057	0.036	0.047
	MLR(RTN)	0.033	0.029	0.028	0.010	0.023
	MLR(CCA-RTN)	0.043	0.036	0.047	0.025	0.036
	Persistence	0.051	0.049	0.047	0.016	0.037
	Random	0.300	0.317	0.280	0.320	0.307
Middle	CCA	0.169	0.205	0.197	0.191	0.196
	MLR(RTN)	0.120	0.136	0.099	0.121	0.121
	MLR(CCA-RTN)	0.133	0.142	0.129	0.146	0.139
	Persistence	0.152	0.143	0.144	0.164	0.151
	Random	0.259	0.254	0.238	0.235	0.245
Low	CCA	0.171	0.180	0.170	0.225	0.190
	MLR(RTN)	0.119	0.139	0.098	0.116	0.120
	MLR(CCA-RTN)	0.125	0.142	0.131	0.134	0.135
	Persistence	0.170	0.219	0.161	0.161	0.181
	Random	0.206	0.211	0.216	0.203	0.209

Table 9. 12-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jul 91 - 1200 UTC 24 Jul 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			20/20 Score (perfect = 1)			
High	CCA	0.805	0.884	0.793	0.882	0.853
	MLR(RTN)	0.834	0.904	0.872	0.964	0.908
	MLR(CCA-RTN)	0.837	0.901	0.821	0.905	0.876
	Persistence	0.835	0.887	0.800	0.938	0.879
	Random	0.259	0.242	0.282	0.242	0.254
Middle	CCA	0.530	0.476	0.486	0.486	0.486
	MLR(RTN)	0.571	0.549	0.586	0.532	0.555
	MLR(CCA-RTN)	0.507	0.471	0.525	0.501	0.497
	Persistence	0.581	0.598	0.562	0.516	0.559
	Random	0.300	0.309	0.327	0.327	0.318
Low	CCA	0.469	0.383	0.504	0.382	0.421
	MLR(RTN)	0.516	0.432	0.560	0.473	0.484
	MLR(CCA-RTN)	0.494	0.389	0.514	0.436	0.444
	Persistence	0.475	0.395	0.488	0.466	0.449
	Random	0.356	0.354	0.347	0.361	0.355
			Normalized Sharpness (perfect = 1)			
High	CCA	1.004	0.988	1.076	0.955	1.000
	MLR(RTN)	0.914	0.982	1.025	1.037	1.003
	MLR(CCA-RTN)	1.004	0.998	1.077	0.975	1.010
	Persistence	0.999	0.999	1.004	1.001	1.001
	Random	0.497	0.468	0.525	0.471	0.485
Middle	CCA	0.918	0.903	1.028	1.070	0.983
	MLR(RTN)	0.837	0.788	0.794	0.792	0.795
	MLR(CCA-RTN)	0.715	0.608	0.737	0.808	0.708
	Persistence	1.008	0.999	0.999	0.998	0.999
	Random	0.587	0.579	0.627	0.656	0.614
Low	CCA	1.068	0.655	1.153	1.132	0.960
	MLR(RTN)	0.752	0.547	0.697	0.468	0.581
	MLR(CCA-RTN)	0.685	0.337	0.828	0.466	0.534
	Persistence	1.006	0.999	0.992	1.003	0.999
	Random	0.779	0.728	0.719	0.791	0.748

In comparing the MAE for the various methods, we find that MLR(RTN) and persistence lead in equally low MAE in January high and middle decks, MLR(RTN) exhibits lowest MAE for January low and July middle and low decks, and persistence is better in the July high deck. MAE for MLR(CCA-RTN) is never lower than MLR(RTN), and is lower than persistence only in the July low deck. The CCA method is competitive only in the high deck for both months.

RMSE and Brier score are similar in that the error penalty increases as the square of the departure of the forecast from the observed. Not surprisingly, they give the same results in the rank ordering of the methods. In all decks in both months, MLR(RTN) has the lowest RMSE and Brier score of any of the forecast methods. In all but the January middle deck, MLR(CCA-RTN) performs better than persistence in RMSE and Brier score. These results are somewhat to be expected because the regression equations are derived on the basis of minimizing mean squared errors. As seen in Table 6, the original MLR(RTN) procedure which used the standard regression slope showed even lower RMSEs than the modified slope method used in these experiments. Persistence gives smaller RMSEs and Brier scores than the CCA method in all but the January high deck.

The 20/20 score gives the fraction of points in which the forecasted cloud amount is within 20 percent of the observed. As in MAE, the rank ordering of skill varies by month and deck. In January, MLR(CCA-RTN) has the highest score for the high deck, and persistence scores best for middle and low decks. In July, MLR(RTN) yields the highest 20/20 scores for the high and low decks, and is virtually tied with persistence in the middle deck. CCA scores the lowest of the four methods in every month-deck category but one, where it is second lowest. Another observation we make about 20/20 scores is that almost without exception, they decrease from high deck through middle deck to low deck for persistence and the MLR methods in both months. In RMSE and Brier scores, this trend is true of persistence but not the MLR procedures. Looking back at MAE, we find that the trend of decreasing skill with decreasing deck altitude is like that of 20/20 score. MAE and 20/20 are similar in that both are a linear measure of agreement with the observed.

Sharpness in a cloud amount field gives the fraction of points in which the cloud amount is within 20 percent of 0 percent or 100 percent. We desire a forecast method which minimizes the loss of sharpness with respect to that of the observed cloud distribution. Thus, we devised normalized sharpness (NS), the ratio of forecast sharpness to observed sharpness as a measure of the loss (NS <1) or gain (NS >1) of sharpness in the forecast process. The NS of persistence is a measure of the constancy of the observed cloud amount distribution between (in this case) 12-hour time periods. It is apparent from the tables that there is essentially no change in the percentage of the total of clear, nearly clear, nearly overcast, and overcast points between consecutive 12-hour interval RTNEPH cloud distributions. Interestingly, the CCA also essentially maintains the sharpness of the observed cloud cover distribution. This is probably because the development of the scheme is based on the frequency of occurrence of the observed cloud amount categories. This fact dictates that in application, the observed cloud amount distribution is preserved. Unfortunately, this fact did not benefit the MLR(CCA-RTN) in regards to sharpness relative to MLR(RTN). In all month-deck categories, MLR(RTN) sharpness is in better agreement with the observed sharpness than MLR(CCA-RTN) sharpness.

We included statistics of the verification of a random forecast of cloud amount as a skill/no skill reference. This would be equivalent to picking a number between 0 and 100 out of a hat for a given gridpoint and time and using that as your cloud amount forecast. The four forecast methods verified in the tables show varying amounts of skill relative to a random selection in all of the skill measures (MAE, RMSE, Brier, 20/20). However, in low cloud decks in both months, the random cloud amount forecast fields (which has by definition an equal portion of the distribution in all 5 percent cloud amount categories) display greater sharpness than forecast fields produced by the MLR procedures. We note that unlike the MLR fields that decrease in NS with decreasing deck altitude, the random fields increase in NS from high deck to low deck. This is because observed sharpness is lower at lower decks in both months (see Figure 8) while the random field sharpness is the same in all decks. However, except for the MLR(CCA-RTN) forecasts in January, the MLR procedures

were not able to capitalize on the lower observed low deck sharpness to improve their NS scores. They, like the RTNEPH fields from which they are derived, exhibit decreasing sharpness with decreasing deck altitude, but at a much faster rate.

In comparing the scores between the two months, we find that for all scores, persistence is a better forecast of high clouds in July, but is a better forecast of middle and low clouds in January. This pattern does not hold for the other three forecast methods. CCA shows better high and low cloud forecast scores for January, and better middle cloud forecast scores for July. MLR(RTN) produces more accurate high cloud forecasts in July, better low cloud scores in January, and about the same scores in both months for middle cloud. MLR(CCA-RTN) scores about the same in all three decks for both months except in 20/20 score, in which high and low cloud forecast scores are better in January and middle clouds have lower scores in July. Finally, the MLR methods produce low deck cloud diagnoses that have considerably higher normalized sharpness in January than July. It is important to state that these trends hold only for this single seven-day verification period. On the basis of a single week, we cannot make any definite conclusions about the seasonal behavior of the various forecast methods.

We now briefly examine the subregional verification statistics for the 12-hour cloud amount forecasts. We will discuss only the most outstanding features seen in comparing the scores in the subregions. First, we observe that when verification takes place over a region that is only a subset of a development region, the bias increases for CCA, MLR(RTN), and MLR(CCA-RTN). As previously stated, we can only be sure that the mean error is minimized only over the development region. CCA suffers the most from this effect, and MLR(CCA-RTN) is affected least of the three. Interestingly, though Table 3 shows a definite trend toward greater middle and low RTNEPH cloudiness over water surfaces than over land surfaces (especially in January), only CCA has a tendency to overpredict over land and underpredict over water. In fact, in January, MLR(RTN) tends to underpredict over land and overpredict over water--a very surprising result. Because we have already stated that subregional application of the statistical forecasting procedures would virtually eliminate subregional biases,

we will discuss this subject no further.

In MAE and 20/20 score, we see a clear pattern of land-water forecast performance characteristics. In both months, with only two exceptions, (CCA middle cloud, 30-90N; July persistence middle cloud; 0-30N) high cloud is better predicted by all methods over water surfaces, while middle and low clouds are most accurately forecast over land. This pattern tends to be true for RMSE and Brier score as well, but with more exceptions. One possible explanation for this pattern may be found in the transformed RTNEPH standard deviation statistics in Table 3. In the middle and low decks in January, the RTNEPH clouds exhibit greater variance over water surfaces than over land, making it harder for forecast procedures to predict accurately over water. However, the greater variance over land in high cloud is apparent only in 30-90N. Furthermore, no clear trend of water vs. land middle and low cloud variance is apparent in July. We cannot conclude, therefore, that regions of higher observed cloud variance will necessarily lead to poorer cloud prediction scores on the basis of our two 1-week samples. We can speculate that the land-water differences in verification scores may suggest a potential for improvement over both surfaces using a subregional application of the statistical forecast methods.

A final outstanding feature of the subregional comparisons is the significant loss of sharpness in low cloud forecasts produced by the MLR procedure over the tropical oceans in both months, and over the extratropical oceans in July. We speculate that the rather homogenous surface evaporation rates over warm oceans in the PL GSM may have led to a more even distribution of cloud amounts over these regions. This speculation is supported by the importance of surface evaporation rate in the MLR procedure as a predictor of low clouds in both months (see Table 7) and the fact that MLR low cloud sharpness is high over cold oceans. Improved surface evaporation parameterization in the model may lead to an improvement in low cloud forecasting over warm oceans.

4.4 12-Hour Cloud Forecast Maps

We now consider a series of regional cloud distribution maps to help us examine the performance of the cloud forecast methods subjectively. In real applications of cloud forecasting methods, it is important that the decision makers have useful and accurate depictions of future cloud distributions in their theater of interest. In the following discussions, we point out the strengths and weaknesses of the subject forecast methods. In all cases, we used the transformed RTNEPH as our standard.

4.4.1 0000 UTC 19 JANUARY 1991 NORTHEASTERN ATLANTIC OCEAN CASE

Figure 24 displays the high cloud deck distribution over the northeastern Atlantic Ocean and western Europe on 0000 UTC 19 January 1991 represented by the transformed RTNEPH and the 12-hour forecast valid at this time and date produced by the three forecast methods: CCA, MLR(RTN), and MLR(CCA-RTN). The most obvious difference between the forecasts and the RTNEPH is the extension of the eastern cloud band southward by the forecasts. The position of this extension coincides with the RTNEPH middle cloud deck distribution as we saw in Figure 18. CCA and MLR(CCA-RTN) produce cloud amounts of >50 percent throughout much of this band, including over the North Sea, where the RTNEPH displays >50 percent cloud amount. In contrast, MLR(RTN) produces >50 percent cloud in this band only at the extreme northern end. This tendency of MLR(RTN) to forecast lesser high cloud amounts is evident also in the more western cloud mass (west of Ireland and the United Kingdom). Here, CCA and MLR(CCA-RTN) include an area of >80 percent cloud cover as is present in RTNEPH, while MLR(RTN) does not. In summary, CCA and MLR(CCA-RTN) are similar in their production of high cloud which equals or exceeds RTNEPH amounts, while MLR(RTN) tends to underpredict high cloud amounts in this case.

In Figure 25, we display the middle cloud distribution for this case. In the middle cloud deck, we find that each forecast method produces unique features. The

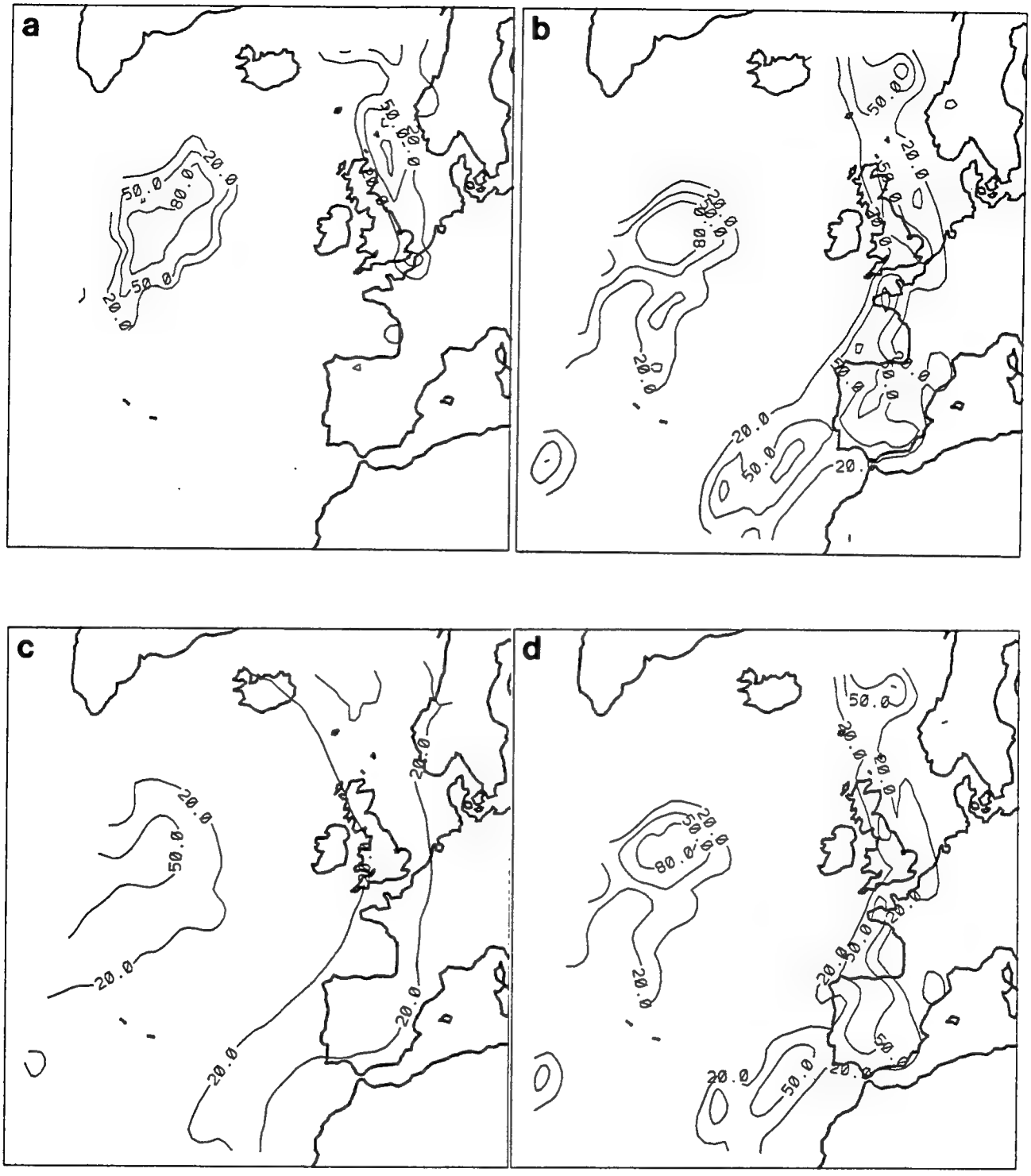

Figure 24. High Deck Cloud Amounts (%) in the Northeastern Atlantic Ocean Region for 0000 UTC 19 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast, (d) 12-Hour MLR(CCA-RTN) Forecast.

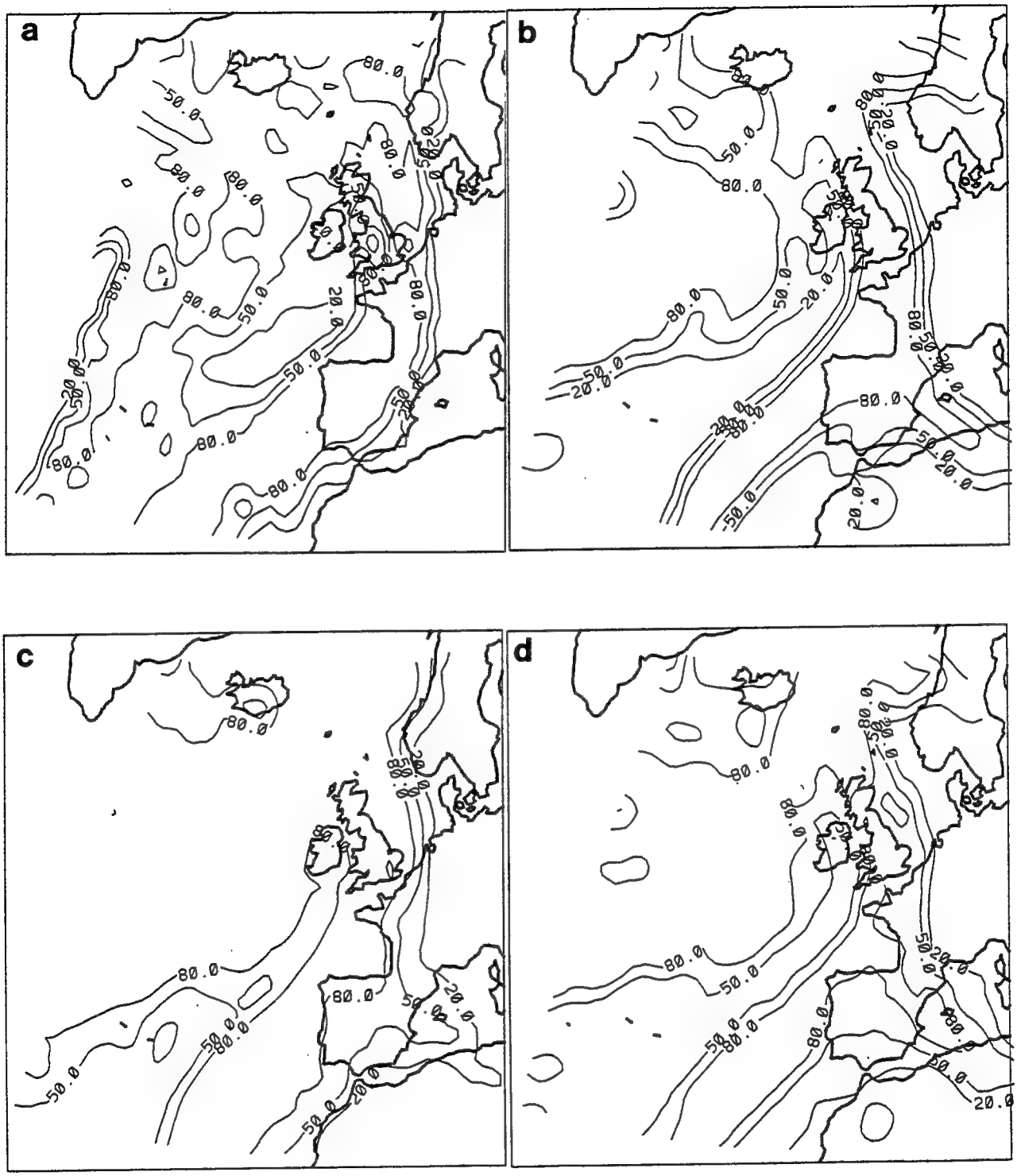

Figure 25. Same as in Figure 24 for Middle Deck Cloud Amounts.

CCA forecast is characterized by extremes of cloud amount in the various sectors. The entire eastern cloud band exceeds 80 percent, the lesser cloud region west of the Iberian peninsula is extensively <20 percent, and the western cloud mass is larger than observed and is made up entirely of cloud amounts >80 percent. As a result of all of this, the CCA cloud amount gradients are sharper than in the RTNEPH. Finally, the CCA forecast produces an apparent spurious extension of middle cloud amount in the Mediterranean Sea and northern Africa. This latter feature is apparent in both MLR forecasts as well. We plotted the middle deck average PL GSM forecast RH and observed this same appendage of high humidities. By contrast with CCA, neither MLR forecast represents the relatively cloud-free region NW of the Iberian peninsula as seen in the RTNEPH. Neither forecast represents the region of between 50 and 80 percent cloud amount NW of the United Kingdom. In general, the two MLR schemes produce too much cloud in the <80 percent areas over the ocean. Over Europe, the position of the leading edge of the cloud band is better represented by the MLR methods than by CCA.

Most of the same tendencies of the forecast schemes in the middle deck are also present in the low deck (Figure 26). CCA produces greater extremes in both lesser and greater cloud amount sectors than RTNEPH, resulting in stronger gradients. This is a reflection of the high CCA NS scores seen in Table 8. MLR(CCA-RTN) and MLR(RTN) both tend to produce too much cloudiness over the ocean. MLR(CCA-RTN) yields much more detail than does MLR(RTN), but neither captures the marine cloud-free region (NW of the Iberian Peninsula) and both tend to produce larger regions with >80 percent cloud cover than are observed. The MLR methods represent low cloudiness over the European continent better than over the oceans.

4.4.2 0000 UTC 23 JULY 1991 NORTHEASTERN ATLANTIC OCEAN CASE

High deck cloud amount maps for a summertime case over exactly the same region are shown in Figure 27. As in the wintertime case just discussed, CCA and MLR(CCA-RTN) produce too much high cloud and MLR(RTN) not enough when

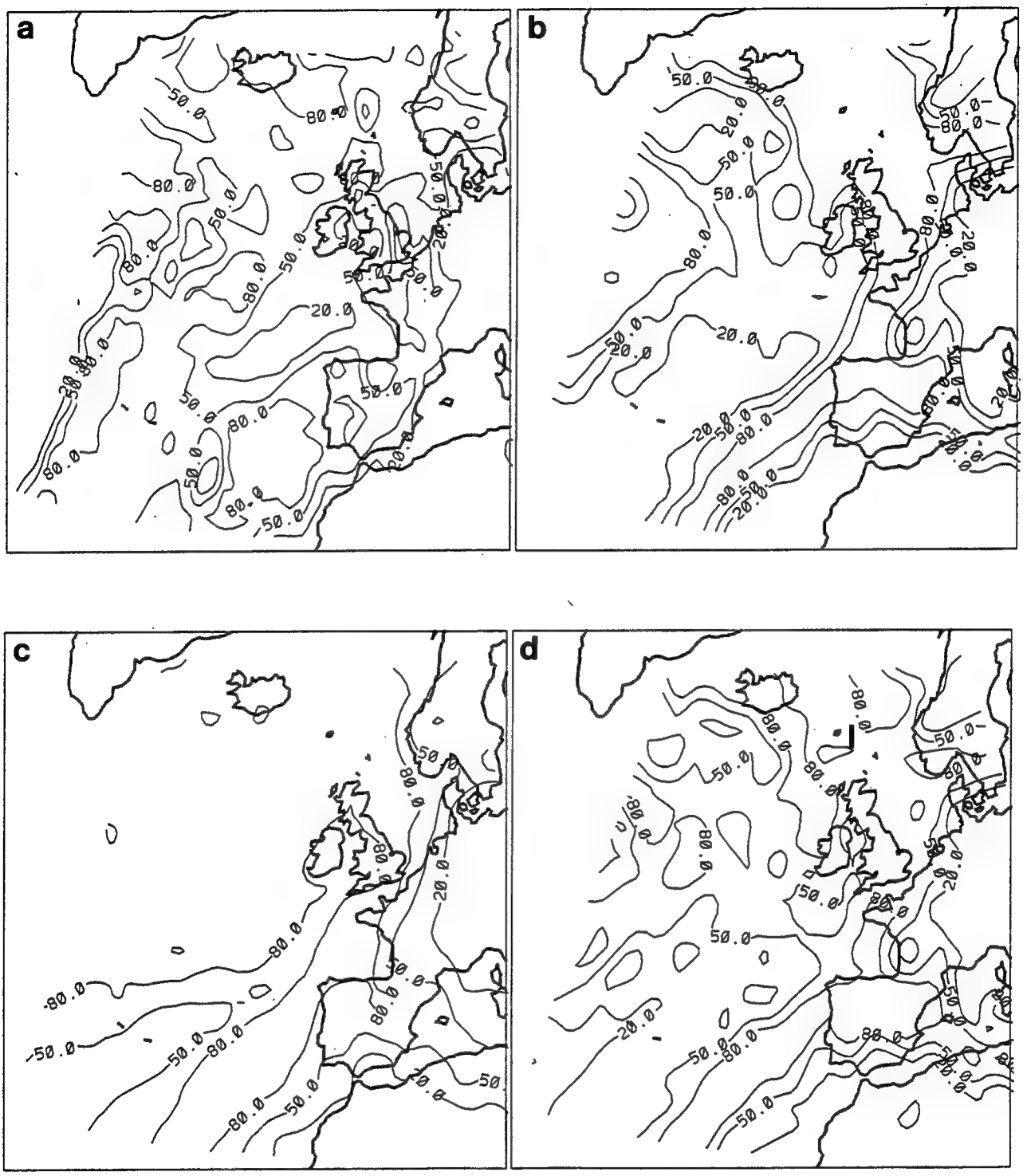

Figure 26. Same as in Figure 24 for Low Deck Cloud Amounts.

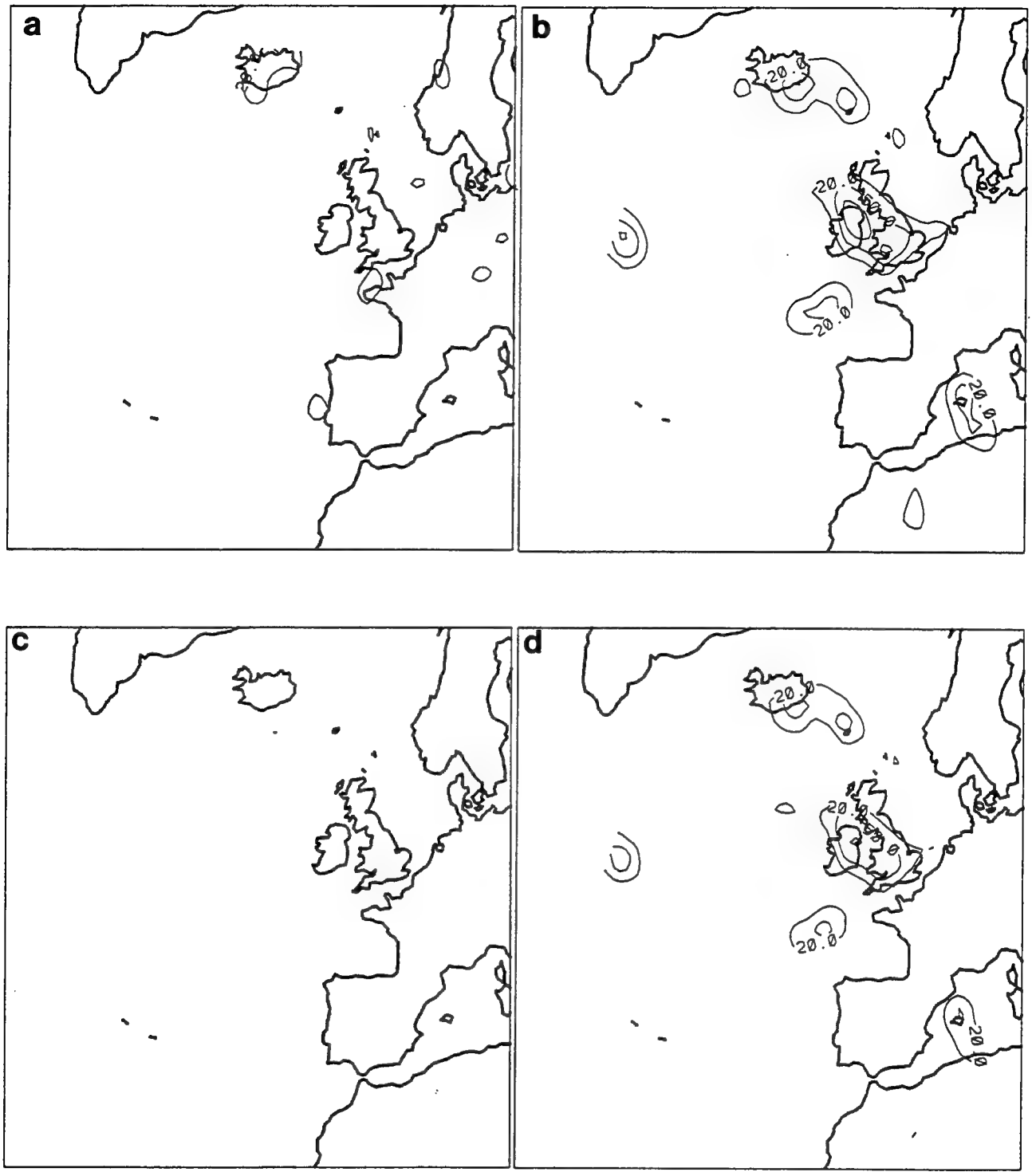

Figure 27. High Deck Cloud Amounts (%) in the Northeastern Atlantic Ocean Region for 0000 UTC 23 July 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast, (d) 12-Hour MLR(CCA-RTN) Forecast.

compared to RTNEPH. In fact, MLR(RTN) produces no cloud amounts in excess of 20 percent in this case. On the other hand, CCA and MLR(CCA-RTN) produce areas of >80 percent cloud cover where less than 20 percent is specified by RTNEPH. Only the cloud mass over Iceland coincides with a similar area of cloudiness in the RTNEPH.

In the middle cloud deck (Figure 28), CCA greatly overextends the area of >80 percent cloud amount (centered west of Ireland and the United Kingdom) relative to RTNEPH. The <20 percent slot SW of the cloud mass is too far SW and too narrow, and RTNEPH's <20 percent slot over the United Kingdom and Ireland are not represented. In addition, CCA creates a major cloud mass over the western Mediterranean Sea and northern Africa that is not represented by RTNEPH. Over the region, CCA produces a positive bias of cloudiness. Characteristically, MLR(RTN) produces flatter gradients than RTNEPH, but in this case, positions the lesser and greater values more correctly than CCA. The <50 percent slot westward of the major cloud mass coincides well with the <20 percent slot in the RTNEPH, and even gives some indication of having <20 percent or slightly more than 20 percent areas in it. Also, the >80 percent area is positioned nearly correctly SE of Iceland, However, the >80 percent band in the MLR(RTN) SW and W of Ireland is too small and located too far to the NE (right over where the RTNEPH has a <20 percent region). MLR(RTN) produces spurious cloudiness over the Mediterranean and northern Africa also, but the amounts are lower than CCA and MLR(CCA-RTN) due to its tendency to underpredict smaller isolated cloud features. Finally, MLR(CCA-RTN) again resembles CCA but with reduced gradients and overall reduced cloudiness. This last fact precludes MLR(CCA-RTN) from properly representing the >80 percent areas in the RTNEPH over the Atlantic.

In the low cloud deck (Figure 29), we see this same tendency of the CCA forecast to produce too much cloud over the ocean. The patterns are similar in the MLR(CCA-RTN) forecasts but the amounts are reduced by about one contour interval (30 percent) in most oceanic locations. The MLR(RTN) scheme produces even less cloudiness, and significantly underpredicts RTNEPH cloud amounts SW and W of Ireland.

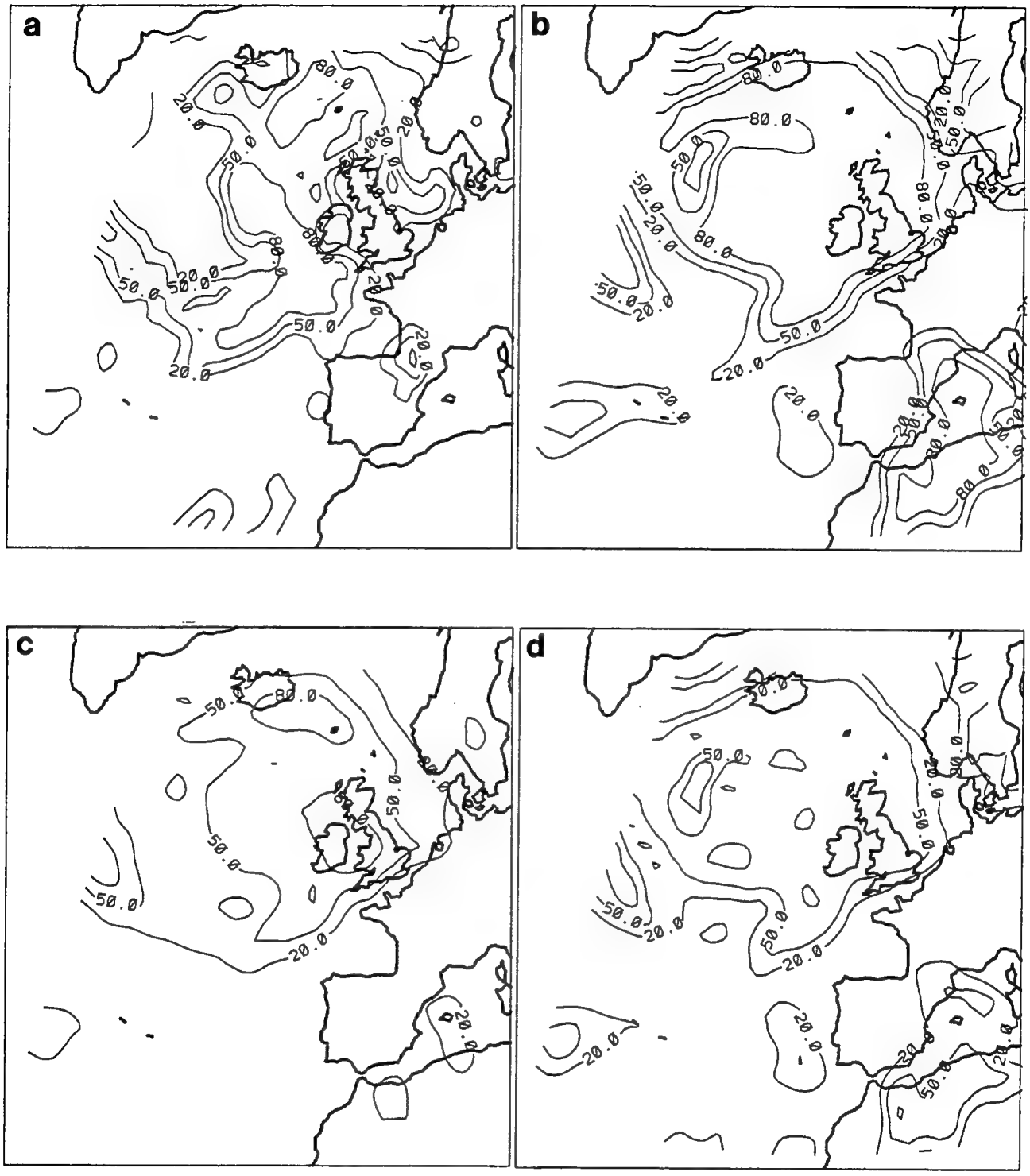

Figure 28. Same as in Figure 27 for Middle Deck Cloud Amounts.

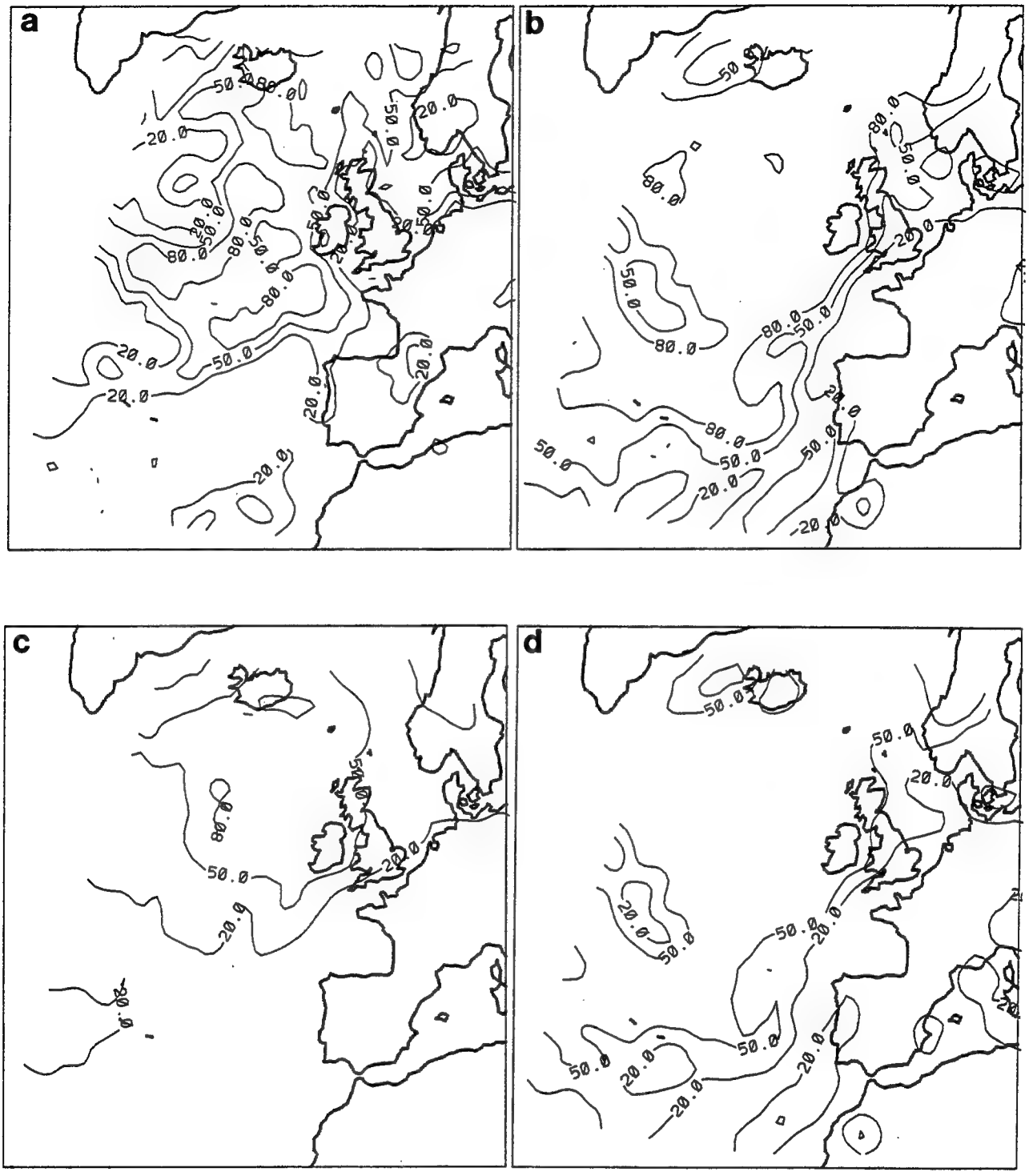

Figure 29. Same as in Figure 27 for Low Deck Cloud Amounts.

In summary, over this region in both seasons, CCA produced too much oceanic cloudiness (especially in summertime) and produced sharper gradients than were observed. The patterns of MLR(CCA-RTN) cloudiness were similar to those of CCA but with reduced cloud amounts and flatter gradients. MLR(RTN) tended to underpredict cloud amounts over the oceans, and had flatter gradients than the other two methods, but positioned the relative maxima and minimum better.

4.4.3 1200 UTC 23 JANUARY 1991 NORTH CENTRAL ASIA CASE

We next consider the 12-hour cloud predictions over an extratropical land surface area for a wintertime case. In contrast to the largely oceanic area discussed in the two previous sections, the present area consists almost entirely of land surface. This region is bounded in longitude by 90E and 130.5E, and in latitude by 40.5N and 69.75N. It was chosen because it was a wholly land surface region that had virtually no missing RTNEPH data and an interesting cloud scene.

In Figure 30, we show the analyzed and predicted high clouds in this region for this case. In all three forecast methods, high cloud amount is underpredicted. This is in contrast to the overprediction by CCA and MLR(CCA-RTN) seen in the wintertime oceanic case described above. The RTNEPH depiction shows that significant amounts of high cloud can be manifest in limited regions, even though the hemispheric average may be underspecified by the transformed RTNEPH. However, the fact that all three forecast methods are developed over the entire hemisphere suggests that the statistical methods may tend to overpredict where few high clouds are present and underpredict in the presence of abundant observed cloudiness. CCA and MLR(CCA-RTN) are once again very similar in their appearance, both essentially missing the most eastern and southern cloud masses. MLR(RTN) smooths out the structure and variability of the cloud masses, but does put the major features in generally the correct location (except for extension up into the NW portion of the region).

In the middle deck (Figure 31), the RTNEPH depiction shows two distinct areas

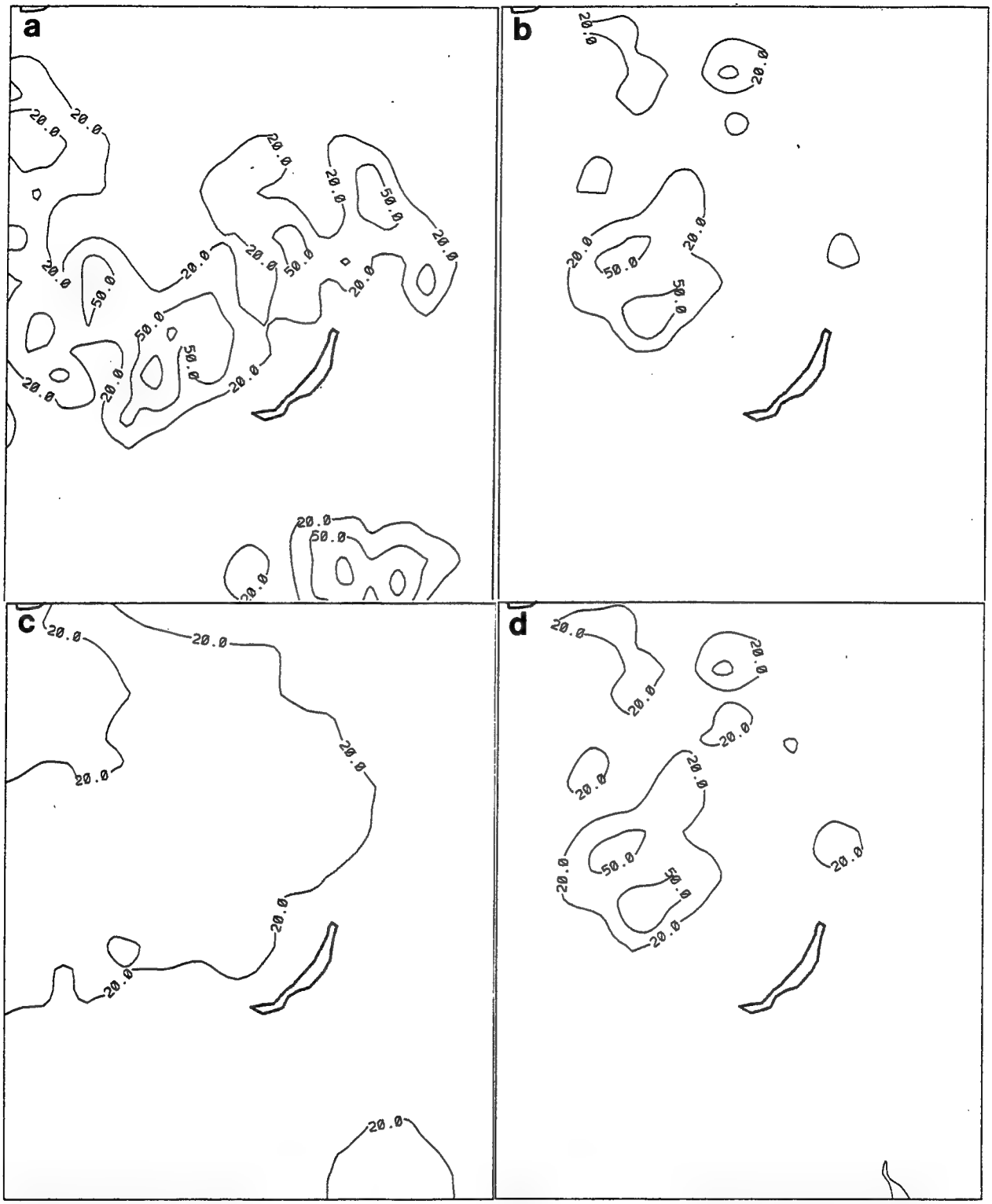

Figure 30. High Deck Cloud Amounts (%) in the North Central Asia Region for 1200 UTC 23 January 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast, (d) 12-Hour MLR(CCA-RTN) Forecast.

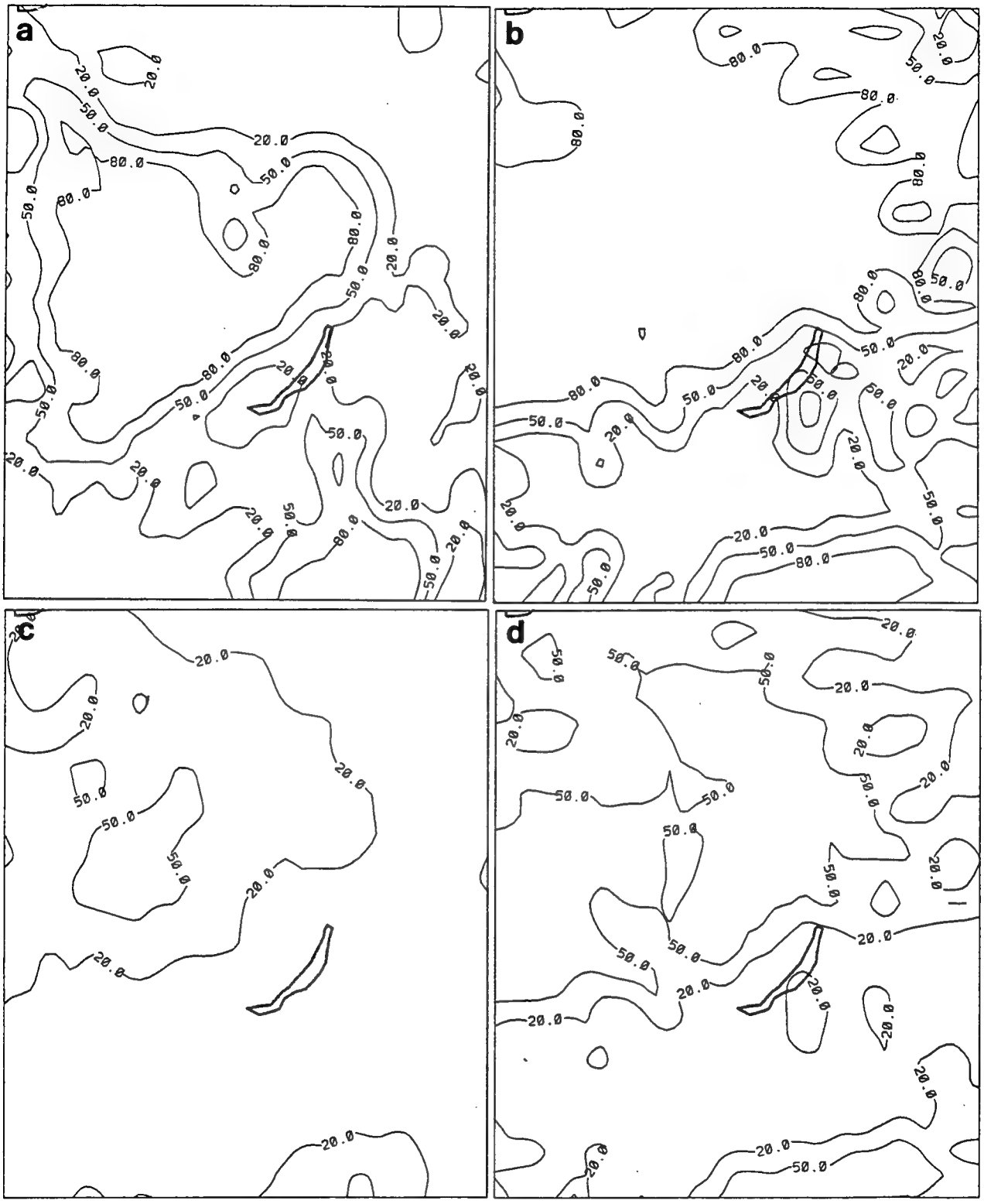

Figure 31. Same as in Figure 30 for Middle Deck Cloud Amounts.

that are nearly overcast separated by a relatively clear slot in between. All three forecast methods replicate this pattern in a relative sense, but with varying errors in the magnitudes. CCA reproduces the southern cloud mass quite well but creates a northern mass that has too large an area in its >80 percent section, and wrongly extends significant cloudiness well into the NE corner. MLR(CCA-RTN) better restricts the areal extent of this cloud mass but underpredicts the amounts in both sectors. This underprediction is even more evident in the MLR(RTN) forecast. In the middle deck, the CCA produces the best overall forecast in this case from a subjective standpoint.

All three forecast methods locate the bulk of the major low cloud mass (Figure 32) too far northward. As in the middle deck, CCA tends to create too large an area of >80 percent cloud amount. MLR(CCA-RTN) once again restricts the size of the area and magnitudes of the densest cloudiness relative to CCA. MLR(RTN) produces an inadequate cloud representation, revealing a cloud distribution that is nearly homogenous and grossly underpredicted. We speculate that this prediction method's dependence on meteorological variables other than RH may have actually hurt it in this case because of the very homogenous snow/ice covered land surface affecting these variables. They were obviously not influenced as much by the synoptic-scale disturbance moving into the area as RH was.

4.4.4 0000 UTC 23 JULY 1991 NORTH CENTRAL ASIA CASE

Figure 33 shows the analysis and forecast high cloud amount distribution over the same region as just discussed, but for a summertime case. As before, the CCA and MLR(CCA-RTN) forecasts look very similar with somewhat greater maxima present in CCA. The RTNEPH cloudiness appears to be less organized than that of the forecasts. The forecasts especially seem to miss the spotty cloudiness in the NE and SW quadrants of the regions.

The transformed RTNEPH depiction of middle clouds (Figure 34) shows a much more organized pattern of cloudiness than in the high cloud. We see a mass of

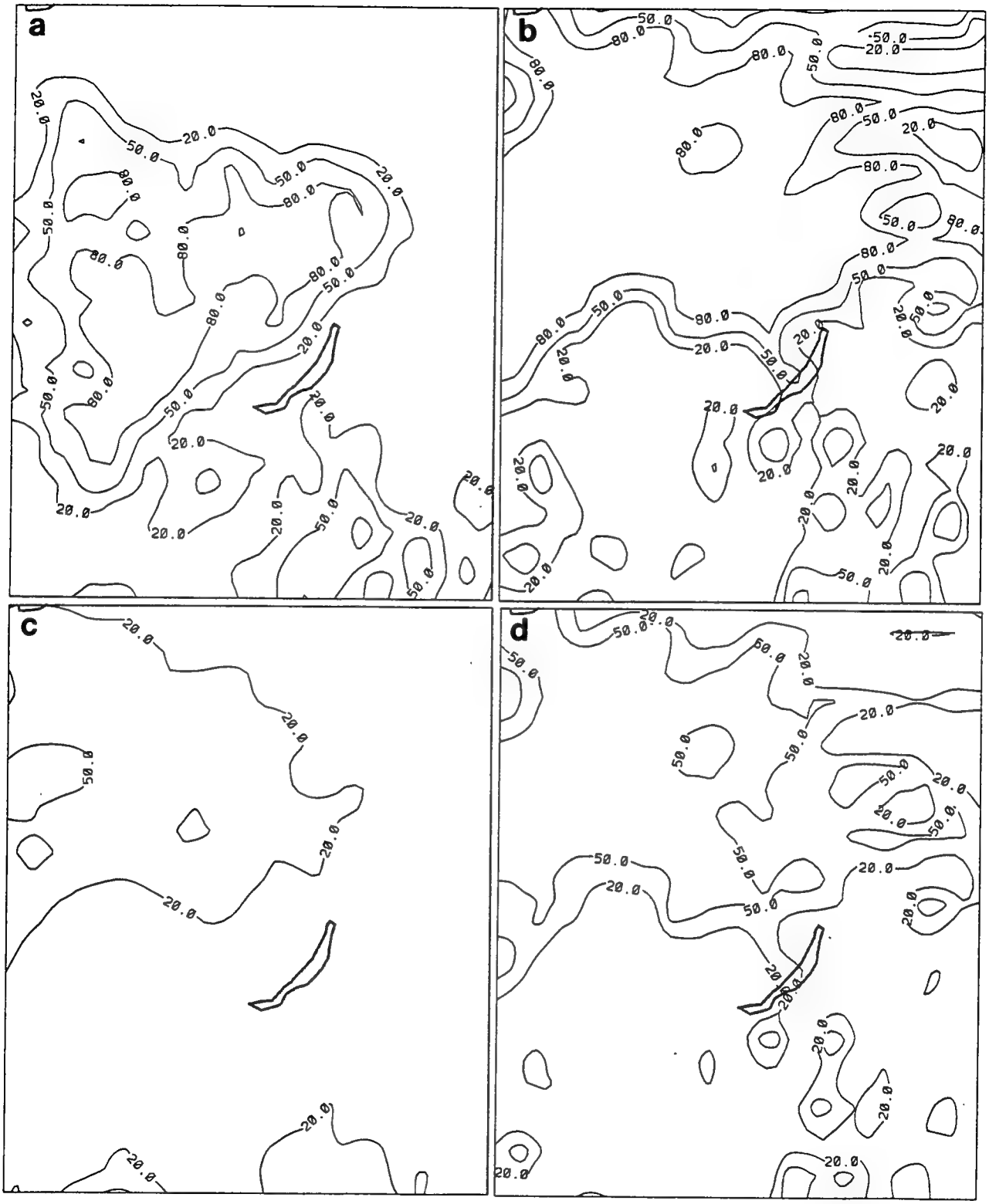

Figure 32. Same as in Figure 30 for Low Deck Cloud Amounts.

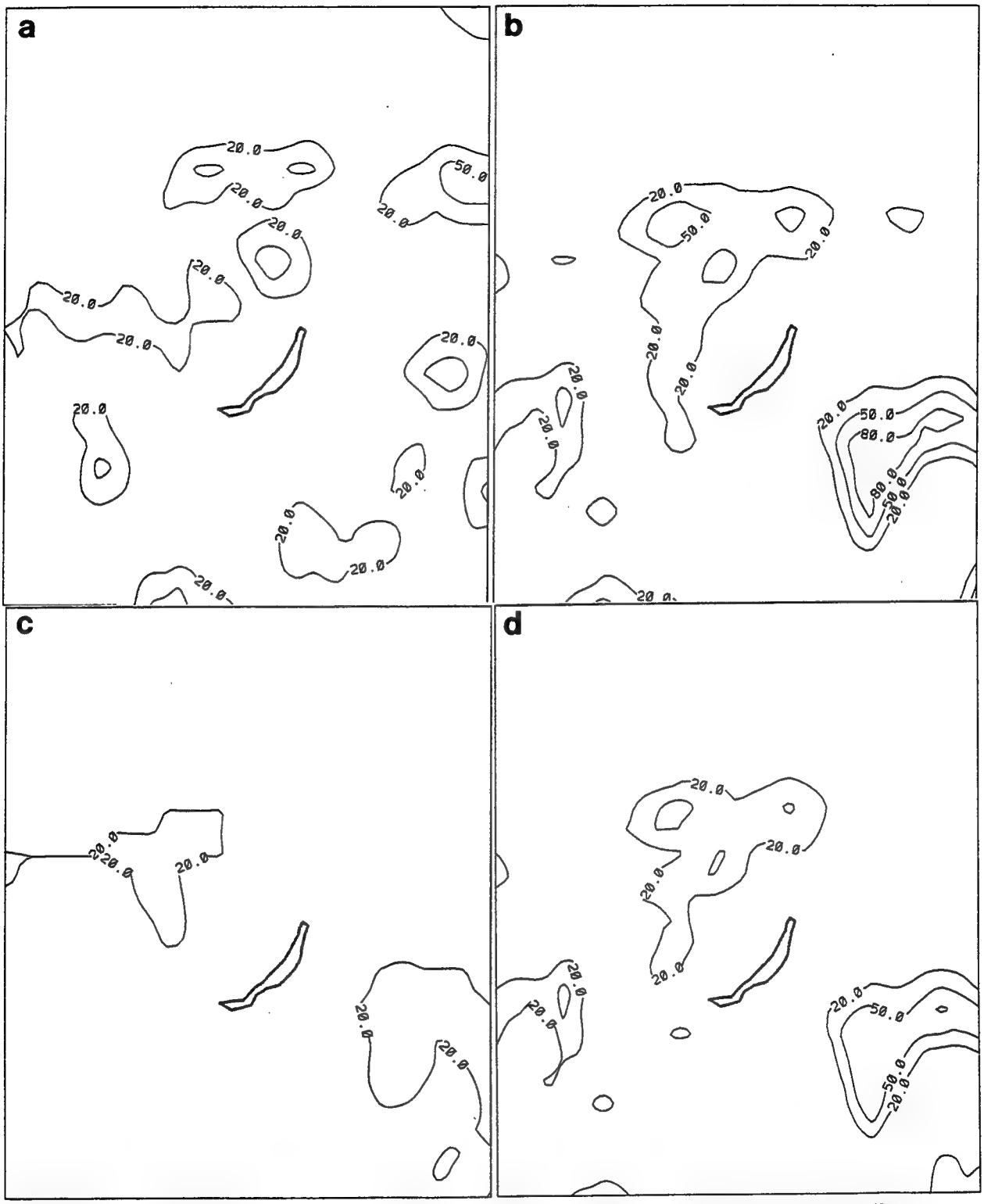

Figure 33. High Deck Cloud Amounts (%) in the North Central Asia Region for 0000 UTC 23 July 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast, (d) 12-Hour MLR(CCA-RTN) Forecast.

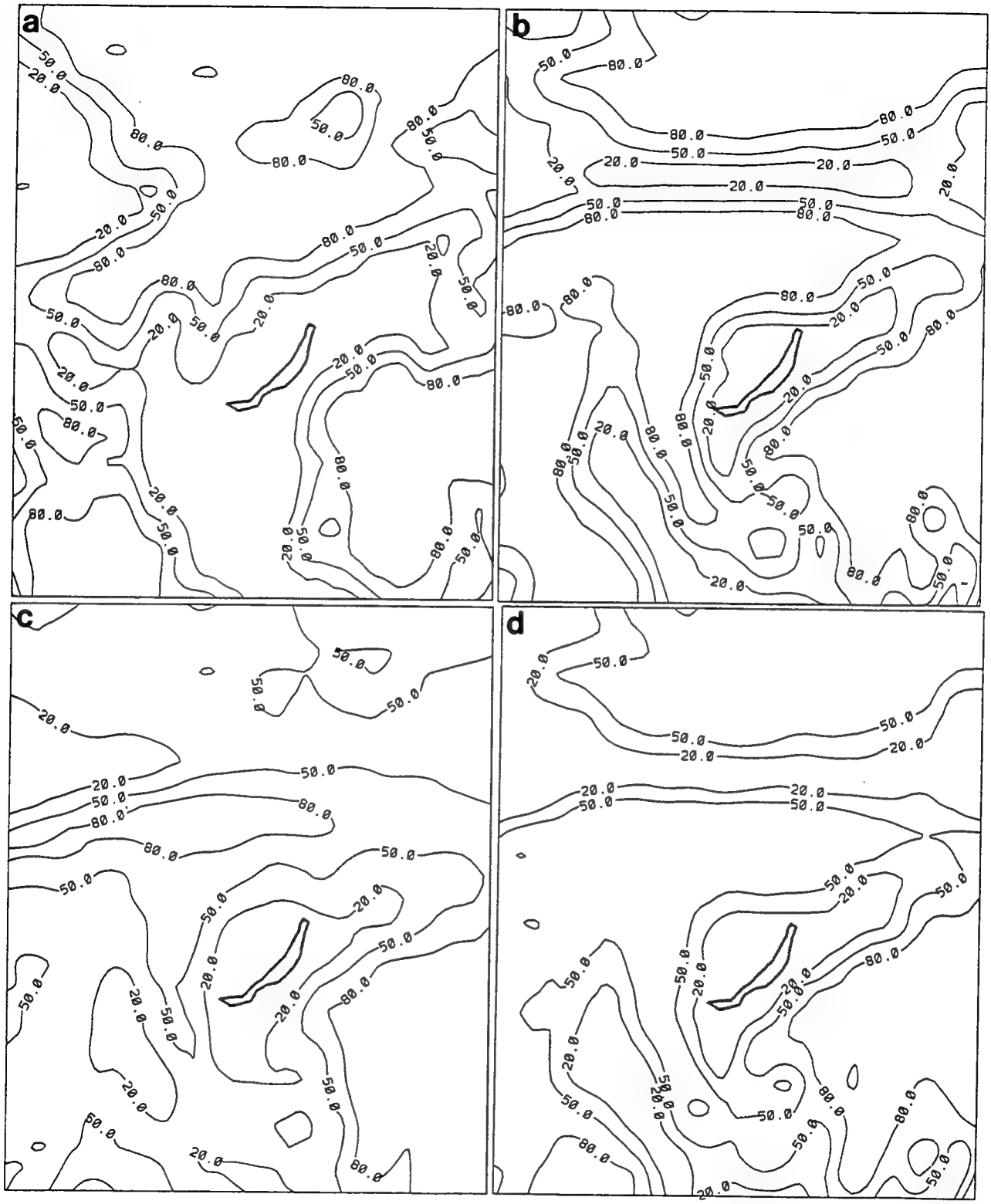

Figure 34. Same as in Figure 33 for Middle Deck Cloud Amounts.

cloudiness in the north central section of the plot, extending southwestward, with smaller areas of >80 percent cloud amount in the SE and SW corners. The CCA forecast reproduced dense cloudiness in the north central sector and in the SE and SW corners. However, a west-east clear band interrupts the north central cloud mass, and the >80 percent mass incorrectly extends southward into the relatively clear area depicted in the RTNEPH. The pattern of the MLR(CCA-RTN) forecast is similar, but with reduced cloud amounts (80 percent > cloud amount >50 percent) in all but the SE sector. The MLR(RTN) also follows the CCA's pattern, but with even less cloudiness in the extreme northern portion and southwestern sector of the plot. On the other hand, MLR(RTN) correctly reintroduces a band of >80 percent cloudiness in the west central portion of the region, and the spurious southward intrusion of cloudiness west of Lake Baikal is reduced in magnitude. As before, the CCA creates the sharpest gradients and has the most area covered by <20 percent and >80 percent cloud cover. In their attempt to reduce the mean squared errors, the MLR techniques create greater areas of cloud amounts between 20 and 80 percent.

These same characteristics are evident in the low cloud distribution shown in Figure 35. However, in this case the RTNEPH reference depicts a great deal more variability and lack of structure than in the middle cloud deck. Therefore, there is less chance that the forecast methods can faithfully reproduce the observed cloud scene. The CCA and MLR(CCA-RTN) are very similar in their forecasts in this case. They capture the major >80 percent cloud sectors present in the RTNEPH--northern, southeastern and west central sectors. They also represent the relatively clear slot over and NE of Lake Baikal. However, they forecast clear areas in a broad swath in the northern half and southwestern part of the plot where RTNEPH shows >50 percent cloudiness, and also forecast 50-80 percent cloudiness in the south central region where the RTNEPH shows clear. MLR(RTN) reduces these deficiencies by producing less cloudiness where RTNEPH is clear and greater amounts of cloudiness where RTNEPH is cloudy. Thus, MLR(RTN) reduces the degree of error from that of the other two techniques at the expense of sharpness in cloud distribution.

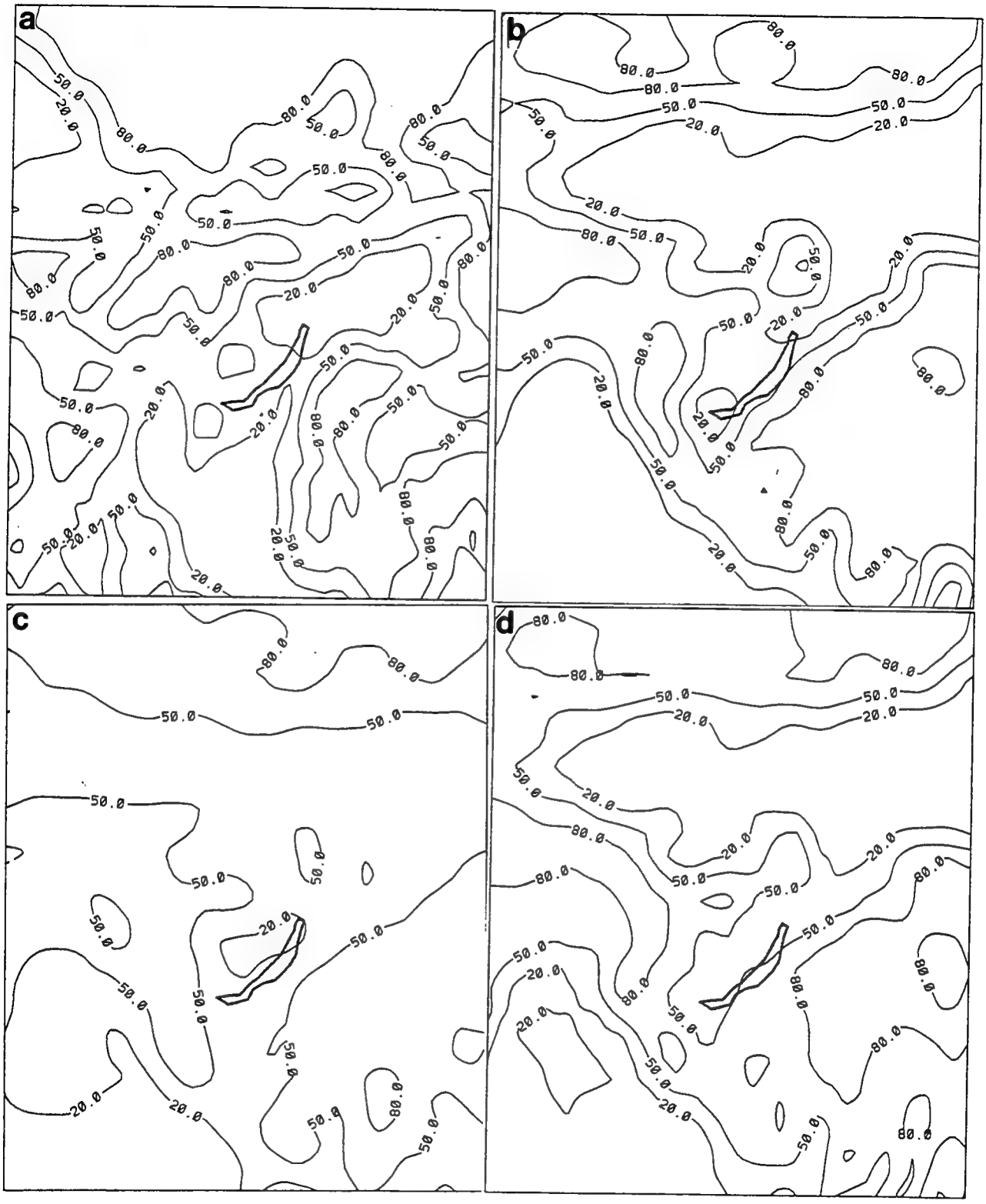

Figure 35. Same as in Figure 33 for Low Deck Cloud Amounts.

4.4.5 0000 UTC 23 JULY 1991 TROPICAL NORTH ATLANTIC OCEAN CASE

We now consider the performance of the three forecast methods in 12-hour forecasts in the tropics. First we focus in on a region centered over the tropical north Atlantic Ocean in a summertime forecast. Climatologically, we expect that intertropical convergence zone (ICZ) cloudiness will dominate the southern portion of our 0-30N latitude extent. We have not included the figures for the high cloud in this case, because the transformed RTNEPH depicted very little high cloudiness for this case. Both CCA and MLR(CCA-RTN) produced significant amounts in the SE and SW corners of the region.

Figure 36 depicts the RTNEPH and 12-hour forecasts of middle deck cloud. Here, for the first time, we see major differences between CCA and MLR(CCA-RTN). The CCA reproduces the basic position of the northern edge of the ICZ quite well in the southern sector of the region, both in location and gradient. MLR(CCA-RTN) tends to reduce the gradient and diffuse the cloud edge, greatly reducing the east-west extent of the >80 percent area. MLR(RTN) better preserves the ICZ shape, cloud edge, and >80 percent area seen in RTNEPH. All three forecasts produce cloudiness in the general locations of the three distinct RTNEPH cloud masses in the NW quadrant of the figures. CCA and MLR(RTN) more accurately simulated the cloud mass containing the >80 percent area closest to the U.S. mainland. MLR(RTN) is most accurate in the forecasts of the two cloud masses in the central Atlantic.

In the RTNEPH rendition of low cloud distribution (Figure 37), a clear zone runs essentially SW to NE across the region, with cloud masses containing primarily 50 to 80 percent cloudiness in the south and NW. None of the forecasts is able to create this pattern. In fact, the CCA produces 50 to 80 percent cloud amounts in the same sector where RTNEPH is clear. The two MLR forecasts lack the distinct cloudy areas in the RTNEPH, creating a rather bland and nondescript cloud field that bears little resemblance to the RTNEPH. We speculate that this may be due to the rather homogenous ocean surface, where spatial variations in predictor variables are reduced.

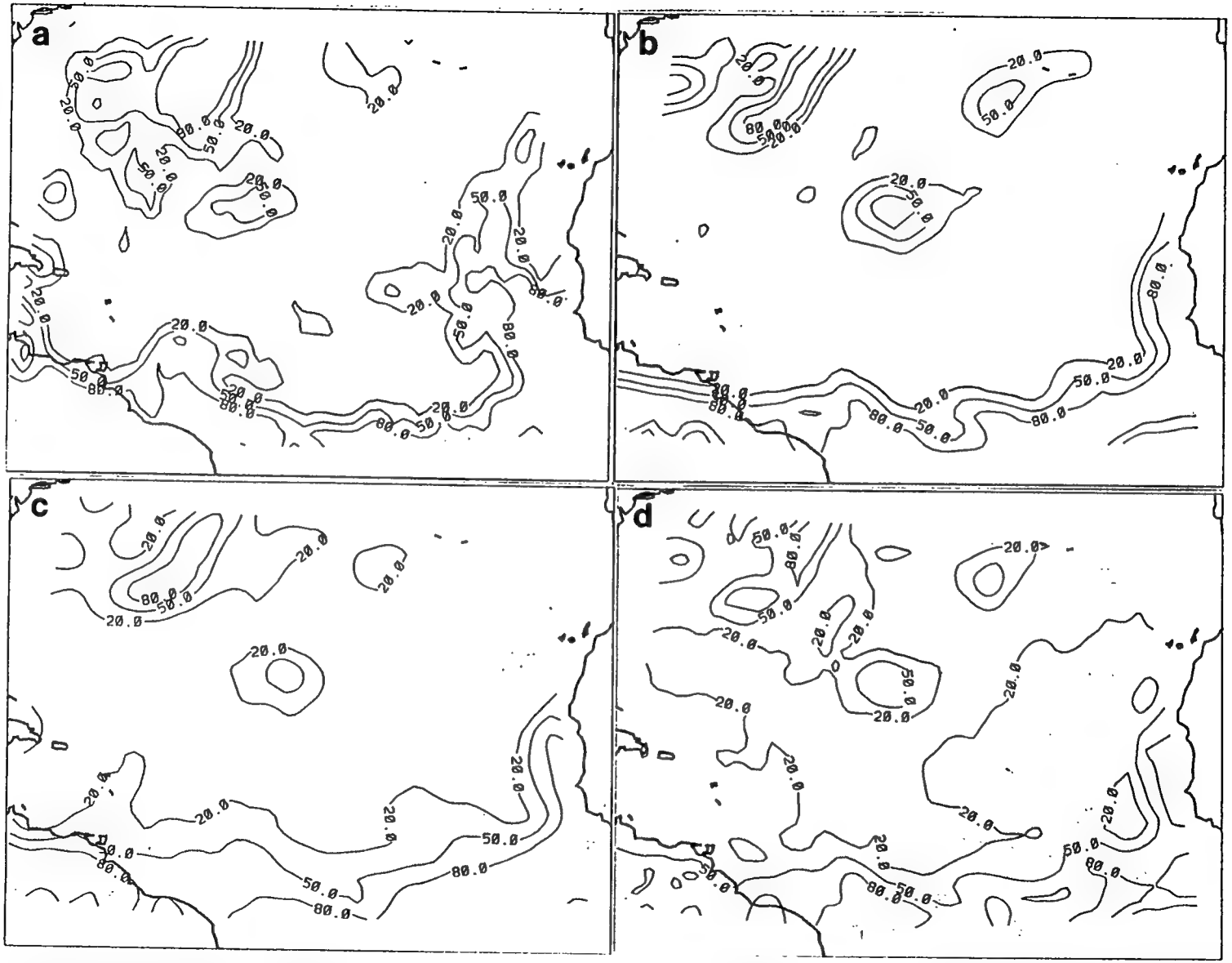

Figure 36. Middle Deck Cloud Amounts (%) in the Tropical North Atlantic Ocean Region for 0000 UTC 23 July 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast, (d) 12-Hour MLR(CCA-RTN) Forecast.

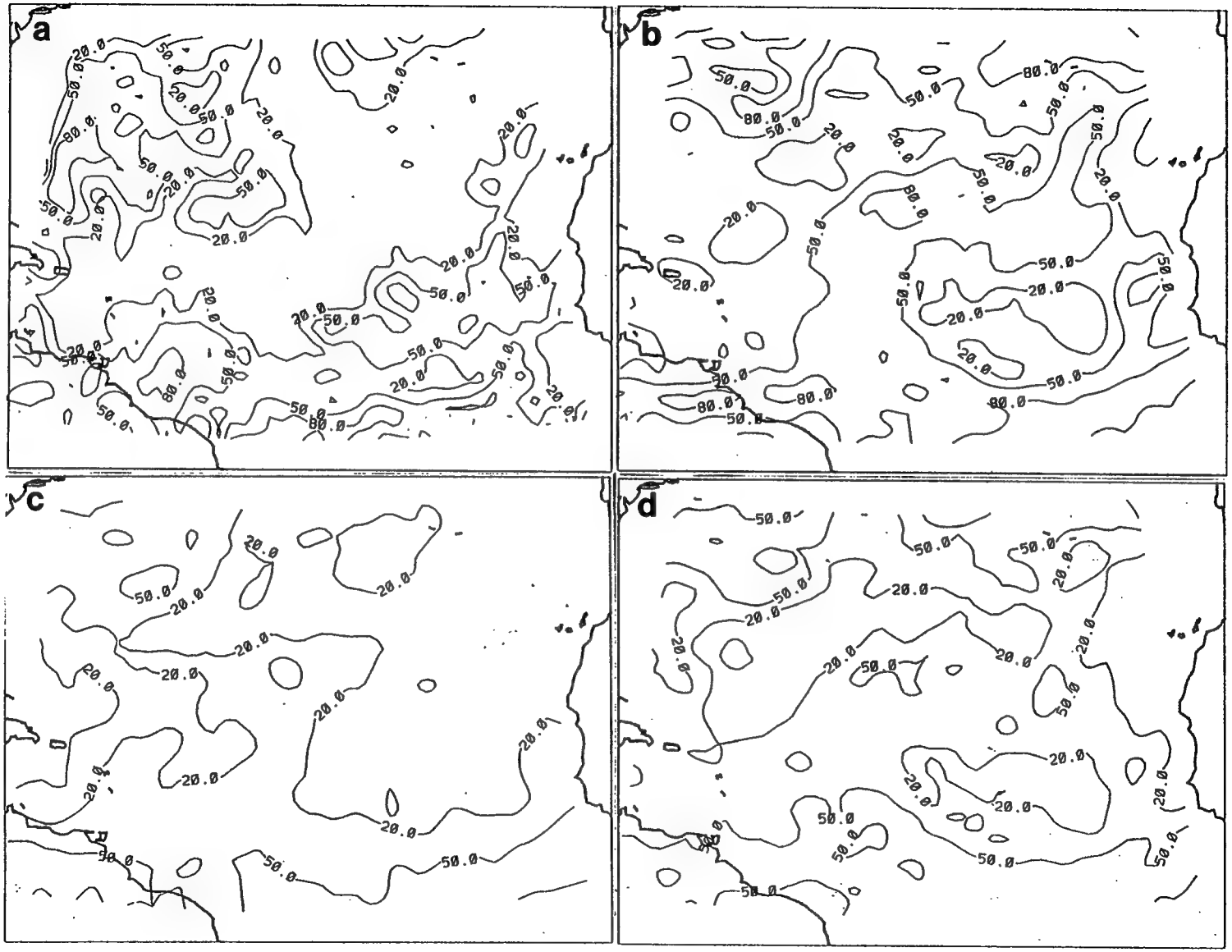

Figure 37. Same as in Figure 36 for Low Deck Cloud Amounts.

4.4.6 0000 UTC 23 JULY 1991 SOUTHEAST ASIA CASE

As the previous case considered largely a tropical ocean region, we now look at the forecast performance of the methods over a tropical land region. In fact, we shall limit our discussion of the cloud distribution maps for southeast Asia shown in Figures 38-40 to just the land areas.

The RTNEPH high cloud depiction shows areas of >20 percent cloudiness over the central SE Asian peninsula, directly to the north of this, and some along the SE China coastline. None of these areas of high cloudiness is produced by any of the forecasts. All three produce a small cloud mass over south central China, which may be a southwestern displacement of the most northerly RTNEPH cloudy region.

In the middle cloud deck, all but the NE sector of the depicted land area is covered by more than 80 percent cloudiness in the RTNEPH. The CCA creates an erroneous gap in the clouds in the SE quadrant of the plot extending over Borneo. The CCA also reduces cloudiness in a band starting at the northern edge of the plot and stretching down over the central SE Asian peninsula, and in the central Malay peninsula. These same patterns of reductions occur in the MLR(CCA-RTN) and MLR(RTN), but to a lesser extent over the SE Asian and Malay peninsulas. Both MLR procedures fail to fully produce the <20 percent region in the NE. In general, the three schemes underforecast the cloud cover in the central SE Asian region in this case.

In low cloud, RTNEPH cloud amounts vary greatly but 20-80 percent predominate. In contrast, CCA produces a large sector of more than 80 percent cloudiness oriented NW to SE. In MLR(CCA-RTN), the >80 percent region is limited to the northern edge, in agreement with RTNEPH.MLR(CCA-RTN) and MLR(RTN) produce a <20 percent area in the NE corner. As usual, MLR(RTN) produces a forecast with less spatial variation, but does cover the SE Asian peninsula with 50-80 percent cloudiness except on the western edge, in agreement with RTNEPH. However, the 50-80 percent cloudiness over Borneo exceeds the amounts shown in the RTNEPH.

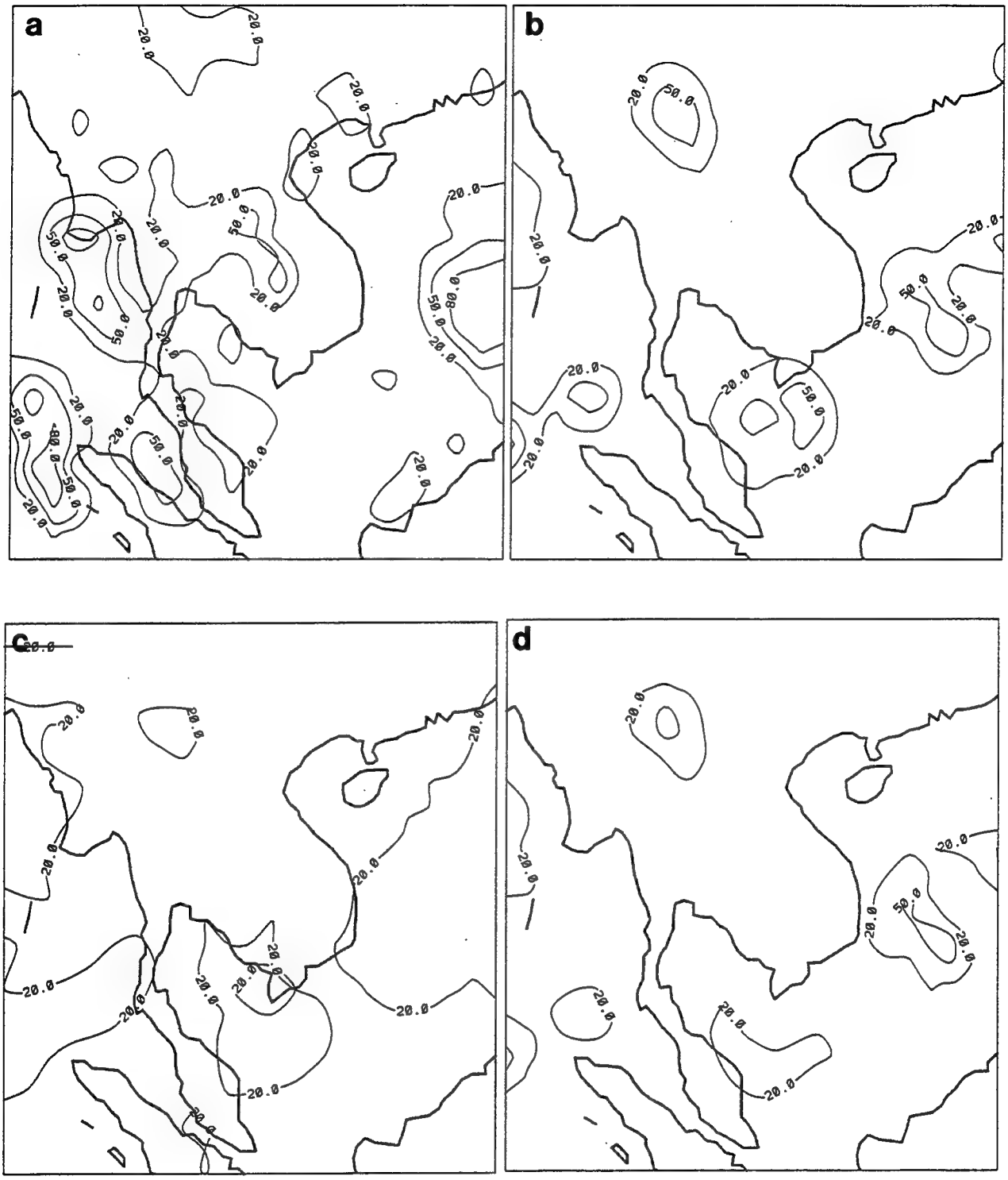

Figure 38. High Deck Cloud Amounts (%) in the Southeast Asia Region for 0000 UTC 23 July 1991 from (a) Transformed RTNEPH, (b) 12-Hour CCA Forecast, (c) 12-Hour MLR(RTN) Forecast, (d) 12-Hour MLR(CCA-RTN) Forecast.

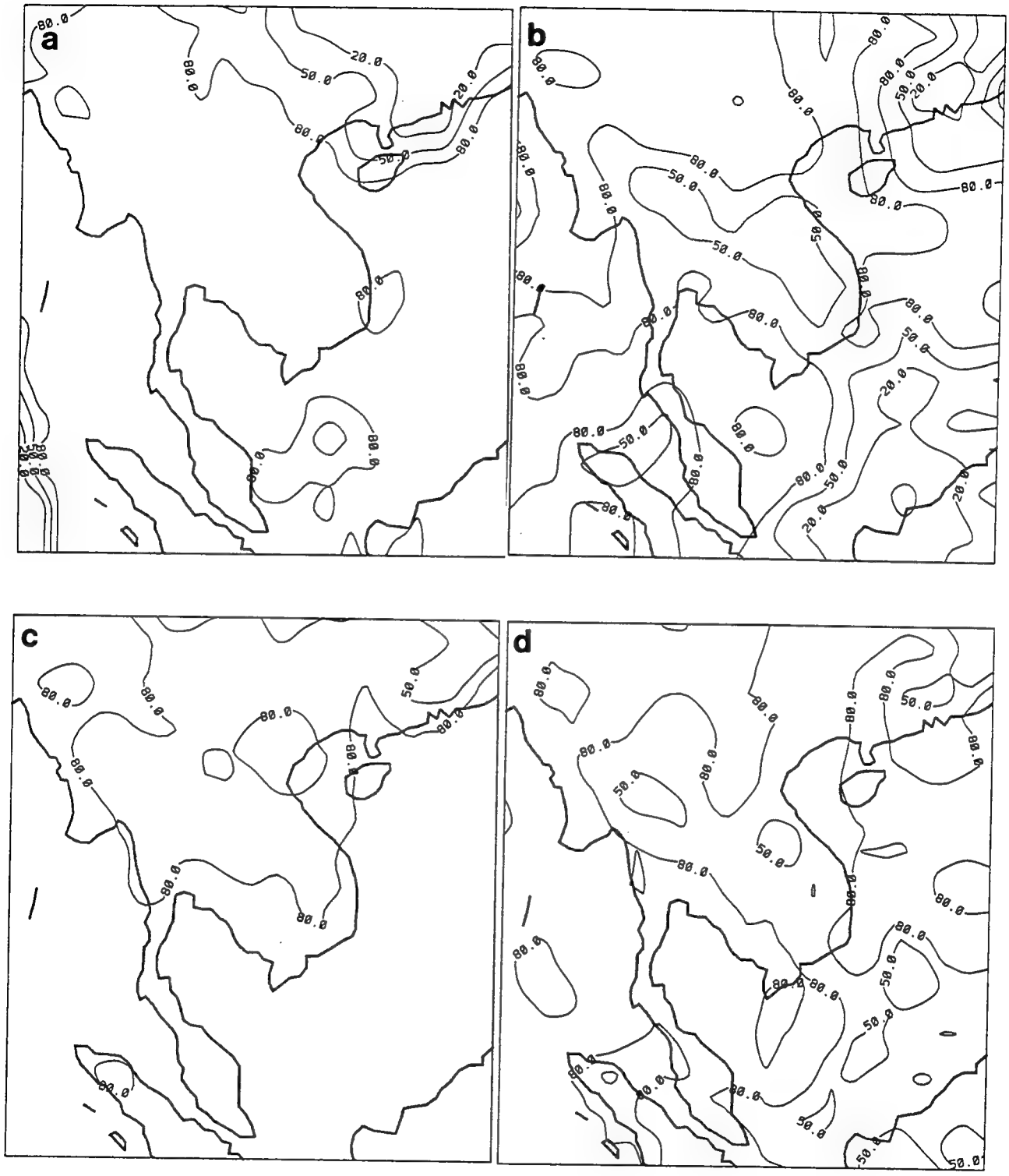

Figure 39. Same as in Figure 38 for Middle Deck Cloud Amounts.

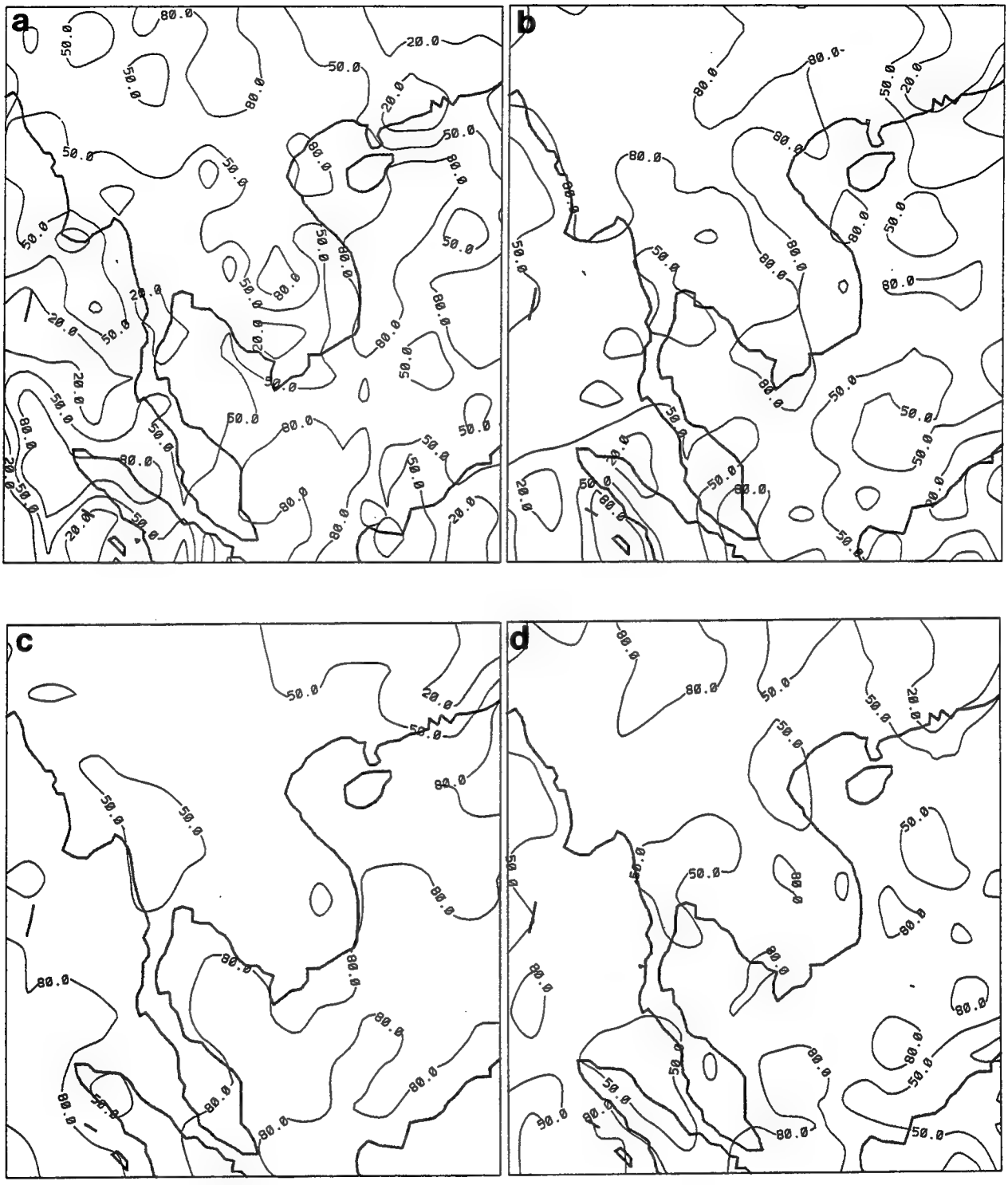

Figure 40. Same as in Figure 38 for Low Deck Cloud Amounts.

4.5 Cloud Amount Predictors for 48-Hour Forecasts

We developed separate predictor-predictand relationships by matching the seven 10-day periods of 48-hour PL GSM forecasts with the transformed RTNEPH cloud amounts valid at forecast times. The resulting seven single-day predictand-predictor relationship sets were then applied to the corresponding verification period day's PL GSM forecast to diagnose the clouds.

We first look at the leading 48-hour forecast predictors as identified by the MLR procedure. In Table 10 we list the strong and useful predictors following the convention of Section 4.2. As in the 12-hour forecasts (Table 7), RH is a strong predictor (except in the low deck in January, where it is only categorized as useful). A further comparison with the leading predictors of the 12-hour forecasts shows a definite reduction in the number of strong and useful dynamic predictors in the 48-hour forecasts. There is a corresponding increase in the number of strong turbulence predictors in January and strong geographic predictors in July. Humidity predictors appear to lose much of their importance in January, while retaining their contribution in July. Over all three decks, the five strongest predictors for the 48-hour forecasts are: January--dry static stability, maximum boundary layer wind speed in a 3X3 array, relative humidity, hours of darkness before valid forecast time t, and predictand deck wind speed; July--relative humidity, sine of latitude, latitude, precipitable water, dry static stability, and evaporation rate.

As in the 12-hour predictors, we sought to determine the degree to which other variables may supplement a single humidity variable as predictors of cloud amount. When 48-hour forecasts are made, the most highly correlated humidity variable and its correlation for January are: low--RH^2 (0.27), middle--RH (0.23), high--RH (0.19); for July: low--condensation pressure spread at t-6 hours (-0.28); middle--precipitable water (0.36), high--condensation pressure spread (0.14). Note that for low and middle decks in both months, the variable was the same as for 12-hour forecasts, but the correlations decreased. For the high deck, the variable changed and the correlation decreased in both months. These humidity variables were the most highly correlated

Table 10. Strong (x) and Useful (+) Predictors in 48-Hour MLR (RTN)

Dynamic	L	M	H	Humidity	L	M	H	Geographic	L	M	H	Turbulence	L	M	H
JANUARY															
temperature			x	relative humidity	+	x	x	latitude		+	+	dry static stab	x	x	x
zonal wind	+			(rel humid)^2	+		+	sin (latitude)		+		moist static stab		x	x
meridional wind		+		precip water		+		cos (latitude)			x	wind speed		x	x
moist convergence			+	cond pres spread		+		sin (longitude)	+			wind shear			+
								cos (zenith angle)	+			BL wind speed			+
								hrs dark before t	x		x	3X3X3 min stab	x	+	
								surface height		+		3X3X3 max shear	x		x
								stnd dev (sfc ht)	+			3X3X3 max speed			+
								% surface water	x			3X3 BL max speed	x	x	x
												evaporation rate			+
												3X3 max conv prcp		x	
JULY															
temperature	x		+	relative humidity	x	x	x	latitude	x	x	+	dry static stab	+	x	x
meridional wind		+	+	precip water	+	x	x	sin (latitude)	x	x	x	moist static stab	x	x	
				cond pres spread	x		+	cos (latitude)	+	x	+	wind shear		+	
								sin (longitude)	+		x	BL wind speed			+
								cos (longitude)	x	x		3X3X3 min stab			x
								cos (zenith angle)	x	+	+	3X3X3 max shear		+	
								hrs dark before t	x	+	x	3X3X3 max speed		x	
								surface height	+			3X3 BL max speed	+		
								% surface water			x	evaporation rate	x	+	x
												3X3 max conv prcp			x
												convective precip			+
												ln (Ri)		+	
												min (sfc ht, stab ht)		+	

of any of the variables except for January low and middle clouds--maximum boundary layer wind in a 3X3 array (0.35 and 0.33 respectively), and July high cloud--sine of longitude (0.19). The leading humidity variables decreased in their correlation from 12- to 48-hour forecasts while non-humidity variables essentially retained their correlation level. This suggests that the growth of the random error in humidity forecasts (error associated with locations and magnitudes of humidity minima and maxima) results in a decrease of correlation with observed cloudiness with forecast duration. This hypothesis is supported by the significant growth of error in the zonal cross sections of RMSE of relative humidity (Figure 15).

4.6 48-Hour Cloud Forecast Verification Statistics

We applied the 48-hour forecast predictor-predictand relationships for the seven 10-day periods to the 48-hour PL GSM forecasts for the corresponding verification days. We then computed the objective verification statistics for the resulting 48-hour forecast cloud diagnoses. Tables 11 and 12 list the scores computed from the 48-hour forecast cloud diagnoses for the complete 7-day verification periods in January and July respectively. In this case, we limit the discussion of the results to the whole hemisphere scores only. We will emphasize the change in scores from the 12-hour forecast diagnoses (Tables 8 and 9) to the 48-hour forecast diagnoses (Tables 11 and 12).

In the January cloud amount forecast scores (compare Table 11 with Table 8), we see no appreciable changes in bias between 12- and 48-hour forecasts. All of the bias values are still small enough to be considered inconsequential. In MAE and 20/20 scores, all procedures show a decrease in skill with forecast duration except MLR(RTN) in high cloud 20/20 score, which shows a slight improvement. Thus, the relative order of skill among the four procedures [CCA,MLR(RTN), MLR(CCA-RTN) persistence) remains about the same for 48 hours as it is for 12 hours. At 48 hours, MLR(RTN) has the lowest MAE for middle and low clouds, whereas persistence is slightly better in high clouds. A strike against CCA and MLR(CCA-RTN) methods is

Table 11. 48-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UT 18 Jan 91-1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			Bias (% Cloud Amount)			
High	CCA	-0.1	-0.1	-3.2	2.2	-0.3
	MLR(RTN)	1.4	-0.2	2.1	2.9	1.5
	MLR(CCA-RTN)	2.0	1.7	0.4	1.7	1.4
	Persistence	-0.4	-0.3	0.4	0.4	0.1
Middle	CCA	-4.1	-15.7	32.9	-6.0	1.2
	MLR(RTN	-9.5	2.7	-7.7	8.5	0.4
	MLR(CCA-RTN)	-3.7	1.6	8.0	-3.8	1.3
	Persistence	-0.9	-1.7	1.6	-1.1	-0.6
Low	CCA	-8.2	-6.0	20.4	-0.6	2.4
	MLR(RTN)	-10.4	5.3	-6.6	9.3	1.7
	MLR(CCA-RTN)	-3.1	4.0	8.4	-1.9	2.8
	Persistence	0.2	-1.2	1.2	-0.2	-0.1
			Mean Absolute Error (% Cloud Amount)			
High	CCA	5.9	4.0	13.3	13.4	9.4
	MLR(RTN)	6.6	3.7	14.0	13.1	9.4
	MLR(CCA-RTN)	7.6	5.6	14.4	13.1	10.3
	Persistence	4.2	3.6	12.6	13.7	9.1
Middle	CCA	22.4	30.2	45.1	40.5	36.3
	MLR(RTN	20.7	27.4	23.9	28.9	26.2
	MLR(CCA-RTN)	23.2	29.3	30.5	33.7	30.3
	Persistence	22.5	32.3	26.0	29.0	28.7
Low	CCA	23.8	32.8	39.6	35.1	34.4
	MLR(RTN)	22.7	28.6	24.7	25.8	26.2
	MLR(CCA-RTN)	24.7	31.5	32.0	30.1	30.6
	Persistence	23.0	33.5	24.6	27.1	28.2

Table 11. 48-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jan 91-1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			RMSE (% Cloud Amount)			
High	CCA	16.4	14.8	26.1	28.2	22.6
	MLR(RTN)	13.0	11.4	20.5	20.5	17.2
	MLR(CCA-RTN)	15.2	13.7	21.4	24.9	20.6
	Persistence	12.6	14.4	24.4	29.7	22.6
Middle	CCA	34.6	41.5	55.7	51.2	47.9
	MLR(RTN	31.7	35.4	34.2	37.8	35.5
	MLR(CCA-RTN)	33.1	36.7	39.1	42.5	38.7
	Persistence	35.9	44.6	39.0	39.7	40.9
Low	CCA	35.6	40.9	52.1	46.1	45.2
	MLR(RTN)	32.0	33.6	34.9	35.1	34.2
	MLR(CCA-RTN)	33.7	37.8	42.4	39.4	39.2
	Persistence	33.7	43.1	37.1	37.0	38.9
			Brier Score (perfect = 0)			
High	CCA	0.027	0.022	0.068	0.079	0.051
	MLR(RTN)	0.016	0.013	0.042	0.042	0.029
	MLR(CCA-RTN)	0.023	0.019	0.059	0.062	0.042
	Persistence	0.016	0.021	0.059	0.088	0.051
Middle	CCA	0.120	0.172	0.310	0.262	0.229
	MLR(RTN	0.101	0.126	0.117	0.143	0.126
	MLR(CCA-RTN)	0.110	0.134	0.153	0.180	0.150
	Persistence	0.129	0.199	0.152	0.158	0.168
Low	CCA	0.127	0.168	0.271	0.212	0.204
	MLR(RTN)	0.103	0.113	0.122	0.123	0.117
	MLR(CCA-RTN)	0.113	0.143	0.180	0.155	0.154
	Persistence	0.114	0.185	0.137	0.137	0.151

Table 11. 48-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jan 91-1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			20/20 Score (perfect = 1)			
High	CCA	0.899	0.940	0.783	0.791	0.851
	MLR(RTN)	0.944	0.967	0.819	0.839	0.889
	MLR(CCA-RTN)	0.910	0.949	0.801	0.813	0.867
	Persistence	0.928	0.948	0.789	0.799	0.859
Middle	CCA	0.627	0.503	0.332	0.378	0.434
	MLR(RTN	0.640	0.492	0.590	0.483	0.530
	MLR(CCA-RTN)	0.612	0.463	0.434	0.417	0.457
	Persistence	0.641	0.494	0.591	0.519	0.541
Low	CCA	0.592	0.415	0.420	0.447	0.443
	MLR(RTN)	0.605	0.397	0.565	0.543	0.504
	MLR(CCA-RTN)	0.572	0.395	0.457	0.479	0.453
	Persistence	0.597	0.432	0.602	0.532	0.523
			Normalized Sharpness(perfect = 1)			
High	CCA	1.007	0.995	1.061	0.971	1.006
	MLR(RTN)	1.043	1.020	1.008	0.948	0.999
	MLR(CCA-RTN)	1.011	0.999	1.056	0.966	1.005
	Persistence	1.008	1.001	0.992	1.002	0.999
Middle	CCA	0.957	0.958	0.895	1.068	0.969
	MLR(RTN	0.916	0.661	0.883	0.870	0.805
	MLR(CCA-RTN)	0.866	0.582	0.561	0.754	0.651
	Persistence	1.011	1.023	0.990	1.010	1.009
Low	CCA	1.189	0.734	1.053	1.131	0.987
	MLR(RTN)	1.004	0.159	0.798	1.081	0.695
	MLR(CCA-RTN)	0.969	0.394	0.728	0.914	0.698
	Persistence	1.005	1.019	0.990	1.013	1.007

Table 12. 48-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jul 91-1200 UTC 24 Jul 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			Bias (% Cloud Amount)			
High	CCA	1.4	-0.5	-2.0	4.6	0.8
	MLR(RTN)	5.5	1.5	3.3	-0.1	1.8
	MLR(CCA-RTN)	3.7	1.2	0.6	3.6	2.3
	Persistence	0.4	0.7	0.5	0.0	0.4
Middle	CCA	0.5	-13.2	14.0	5.7	0.7
	MLR(RTN	1.7	7.1	-5.6	-0.9	1.0
	MLR(CCA-RTN)	0.9	0.1	8.0	2.1	1.3
	Persistence	-1.7	-0.2	-1.8	-0.7	-0.9
Low	CCA	-1.2	-1.3	-10.5	19.9	2.8
	MLR(RTN)	3.5	7.2	-4.6	0.2	1.8
	MLR(CCA-RTN)	1.2	-0.2	-2.7	4.2	0.6
	Persistence	-0.9	-0.3	-1.0	0.1	-0.4
			Mean Absolute Error (% Cloud Amount)			
High	CCA	12.0	6.9	12.8	8.5	9.3
	MLR(RTN)	12.1	7.8	12.3	4.1	8.2
	MLR(CCA-RTN)	12.5	9.1	12.9	7.5	9.9
	Persistence	8.7	7.5	10.1	4.2	7.1
Middle	CCA	28.6	33.0	36.5	37.4	34.9
	MLR(RTN	23.8	28.5	26.1	29.4	27.7
	MLR(CCA-RTN)	26.9	30.6	30.8	33.9	31.4
	Persistence	28.0	33.2	28.1	34.5	32.0
Low	CCA	31.5	35.9	31.1	39.1	35.3
	MLR(RTN)	27.0	30.3	24.7	27.7	27.8
	MLR(CCA-RTN)	28.3	32.1	28.4	32.5	31.0
	Persistence	29.7	34.7	27.2	31.4	31.4

Table 12. 48-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jul 91-1200 UTC 24 Jul 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			RMSE (% Cloud Amount)			
High	CCA	24.9	19.6	24.7	20.7	21.8
	MLR(RTN)	18.1	16.1	17.4	10.1	15.1
	MLR(CCA-RTN)	21.0	18.1	22.0	16.6	19.0
	Persistence	20.3	21.8	19.8	13.1	18.7
Middle	CCA	40.8	44.5	48.5	49.2	46.7
	MLR(RTN	34.7	38.7	34.9	37.7	37.1
	MLR(CCA-RTN)	37.2	39.6	40.4	43.4	40.8
	Persistence	41.2	46.5	39.9	46.0	44.5
Low	CCA	41.5	43.9	42.3	48.9	44.9
	MLR(RTN)	35.8	37.3	32.1	34.0	34.9
	MLR(CCA-RTN)	36.1	38.2	36.9	39.8	38.2
	Persistence	39.4	44.4	37.4	40.8	41.2
			Brier Score (perfect = 0)			
High	CCA	0.062	0.038	0.061	0.043	0.048
	MLR(RTN)	0.033	0.026	0.030	0.010	0.023
	MLR(CCA-RTN)	0.044	0.033	0.048	0.028	0.036
	Persistence	0.041	0.048	0.039	0.017	0.035
Middle	CCA	0.167	0.198	0.235	0.242	0.218
	MLR(RTN	0.120	0.150	0.122	0.142	0.138
	MLR(CCA-RTN)	0.138	0.157	0.163	0.189	0.166
	Persistence	0.170	0.217	0.159	0.211	0.198
Low	CCA	0.172	0.192	0.179	0.239	0.201
	MLR(RTN)	0.128	0.139	0.103	0.115	0.122
	MLR(CCA-RTN)	0.130	0.146	0.136	0.159	0.146
	Persistence	0.155	0.198	0.140	0.166	0.170

Table 12. 48-Hour Cloud Amount Forecast Verification Scores for Initial Times of 0000 UTC 18 Jan 91-1200 UTC 24 Jan 91

Deck	Method	0-30N Land	0-30N Water	30-90N Land	30-90N Water	0-90N
			20/20 Score (perfect = 1)			
High	CCA	0.797	0.892	0.785	0.862	0.848
	MLR(RTN)	0.827	0.915	0.846	0.965	0.905
	MLR(CCA-RTN)	0.825	0.904	0.815	0.890	0.870
	Persistence	0.855	0.893	0.827	0.936	0.889
Middle	CCA	0.545	0.480	0.444	0.430	0.461
	MLR(RTN	0.585	0.503	0.590	0.461	0.501
	MLR(CCA-RTN)	0.550	0.467	0.468	0.433	0.464
	Persistence	0.554	0.500	0.548	0.454	0.500
Low	CCA	0.460	0.368	0.490	0.366	0.406
	MLR(RTN)	0.499	0.424	0.531	0.455	0.467
	MLR(CCA-RTN)	0.480	0.369	0.497	0.406	0.422
	Persistence	0.477	0.416	0.533	0.455	0.462
			Normalized Sharpness(perfect = 1)			
High	CCA	0.987	0.986	1.062	0.940	0.990
	MLR(RTN)	0.912	0.989	0.984	1.038	0.997
	MLR(CCA-RTN)	0.978	0.989	1.061	0.955	0.994
	Persistence	0.999	0.994	0.989	1.000	0.995
0						
Middle	CCA	0.988	0.907	1.040	1.091	1.001
	MLR(RTN	0.860	0.748	0.721	0.631	0.717
	MLR(CCA-RTN)	0.862	0.673	0.747	0.790	0.742
	Persistence	1.009	1.005	1.006	0.998	1.003
Low	CCA	1.091	0.670	1.199	1.170	0.990
	MLR(RTN)	0.751	0.497	0.676	0.385	0.533
	MLR(CCA-RTN)	0.640	0.272	0.805	0.562	0.528
	Persistence	1.002	0.999	0.988	1.007	0.999

that persistence still has lower MAE at 48 hours. This is true of CCA and both MLR methods in 20/20 score in middle and low cloud.

In measures of mean-squared error (RMSE and Brier), the relative order of skill changes from 12 to 48 hours between MLR(CCA-RTN) and persistence. MLR(CCA-RTN) has the second best and persistence the third best scores at 12 hours, but these positions are reversed at 48 hours. MLR(RTN) holds the position of having the best RMSE and Brier scores in all decks for both forecast times. In fact, MLR(RTN) produces the smallest percentage increase of RMSE and Brier from 12 hours to 48 hours of any of the four methods in all three decks except RMSE in high cloud, where the increase is about the same as for persistence. This fact suggests that MLR(RTN) maintains its skill with increasing forecast time better than the other methods.

All four methods maintain sharpness of the cloud distribution in the high deck. However, in the middle deck, CCA and both MLR(RTN) yield a slightly degraded sharpness in the 48-hour forecasts. In the low deck, MLR(RTN) and MLR(CCA-RTN) produce less sharp cloud distributions than for 12-hour forecasts. In the middle and low decks, the MLR methods consistently produce cloud distributions with lower sharpness than do CCA and persistence.

July cloud amount bias scores in Table 12 remain inconsequentially low as they did in the January 48-hour forecast diagnoses. MAE and 20/20 skill generally degrade with forecast time for the four procedures, with a few minor exceptions. Persistence remains most skillful in MAE in high cloud, while MLR(RTN) continues to have lowest MAE in middle and low decks. MLR(RTN) is as good as or better than persistence in 20/20 score in all decks in 48-hour forecast diagnoses.

July 48-hour RMSE and Brier scores in the high cloud deck reflect no loss of skill for any of the methods from the 12-hour values. Similarly, persistence in the low deck is slightly more skillful at 48 hours than at 12 hours. Once again, MLR(RTN) produces the lowest mean-squared errors at both forecast times in all three decks. The increase in mean-squared error from the 12-hour to 48-hour diagnoses is largest in the middle deck. Yet even in this case, MLR(RTN) maintains its skill better than two of the three other procedures.

As in January, sharpness is essentially maintained in the high cloud deck for all four procedures in the July diagnoses. In the middle deck, MLR(CCA-RTN) diagnoses increase slightly in sharpness between 12 and 48 hours, but sharpness is lost in the middle deck by MLR(RTN) and by both MLR methods in the low deck. As in the 12-hour diagnoses, MLR methods lose normalized sharpness with decreasing deck altitude even though the reference (RTNEPH) sharpness is less in the lower decks.

4.7 48-Hour Cloud Forecast Maps

We also produced maps of cloud amounts diagnosed from 48-hour PL GSM forecasts. Our discussion of these maps is limited to two extratropical land cases (summer and winter) and a tropical land case (summer only). The focus of these discussions will be a comparison of the 48-hour forecast diagnoses with the 12-hour forecast diagnoses discussed earlier.

4.7.1 1200 UTC 23 JANUARY 1991 NORTH CENTRAL ASIA CASE

The cloud scenes discussed here were the product of 48-hour diagnoses conducted at the same time and region discussed in Section 4.4.3. Thus, the 48-hour forecast depictions (Figures 41-43) can be compared directly with the corresponding 12-hour forecast maps (Figures 30-32).

In the high cloud deck (compare Figures 30 and 41), the CCA and MLR(CCA-RTN) cloud distributions are much more scattered at longer forecast times. In the MLR(RTN) depiction, the indicated 48-hour forecast cloudiness is spread eastward in the northern half of the region in comparison to the 12-hour forecasts. Clouds in the extreme southern part of the region are not singly oriented in the location of greatest RTNEPH cloudiness as they are at 12 hours.

CCA and MLR(CCA-RTN) middle cloud distributions (compare Figures 31 and 42) produce their major areas of greatest and least amount of cloud in generally the same locations at both forecast times. Again, 48-hour forecasts depict a greater degree

of variance, especially in relatively clear area in the SW portions of the region. The CCA 48-hour forecast is especially prone to small "tongues" of >80 percent cloudiness in what is a <20 percent area at 12 hours. This no doubt contributes to the poorer RMSE scores of the 48-hour CCA forecasts. By contrast, the 48-hour MLR(RTN) forecast produces less variance by removing most of the >50 percent areas evident at 12 hours. This also leads to an increase RMSE, but not as much as in CCA.

CCA and MLR(CCA-RTN) low cloud depictions (compare Figures 32 and 43) show even more of the evidence short wavelength variance at 48 hours than was present at 12 hours. From the CCA figures, it appears that some phenomenon in the forecast excited a great deal of spatial irregularity in the RH field, leading to a very patchy cloud distribution. The intensity of the variance increases from 12 to 48 hours. By contrast, the MLR(RTN) cloud distribution maintains a low degree of cloud amount variance, but loses some of the > 50 percent cloudiness at 48 hours.

4.7.2 0000 UTC 23 JULY 1991 NORTH CENTRAL ASIA CASE

We diagnosed cloud distributions from 48-hour forecasts valid at the same time and over the same location as the 12-hour diagnoses discussed in Section 4.4.4. In the following discussion, we compare the 48-hour forecast cloud diagnoses (Figures 44-46) with corresponding depictions of the 12-hour forecast maps (Figures 33-35).

In comparing high cloud deck cloud amount distributions in Figures 33 and 44, we find that the CCA and MLR(CCA-RTN) diagnoses resemble each other more at both forecast times than do the 12- and 48-hour diagnoses from a given scheme. This illustrates the great degree of influence that the CCA has on MLR(CCA-RTN) diagnoses. At 12 hours and especially at 48 hours, the MLR(CCA-RTN) diagnoses less cloud in most of cloud masses than does CCA. Unfortunately, the result is that none of the four figures resembles the transformed RTNEPH cloud distribution. Further, the MLR(RTN) agrees more with the transformed RTNEPH than do CCA and MLR(CCA-RTN) only because it produces very little cloud. As in the January case, the 48-hour amounts are less than the 12-hour amounts.

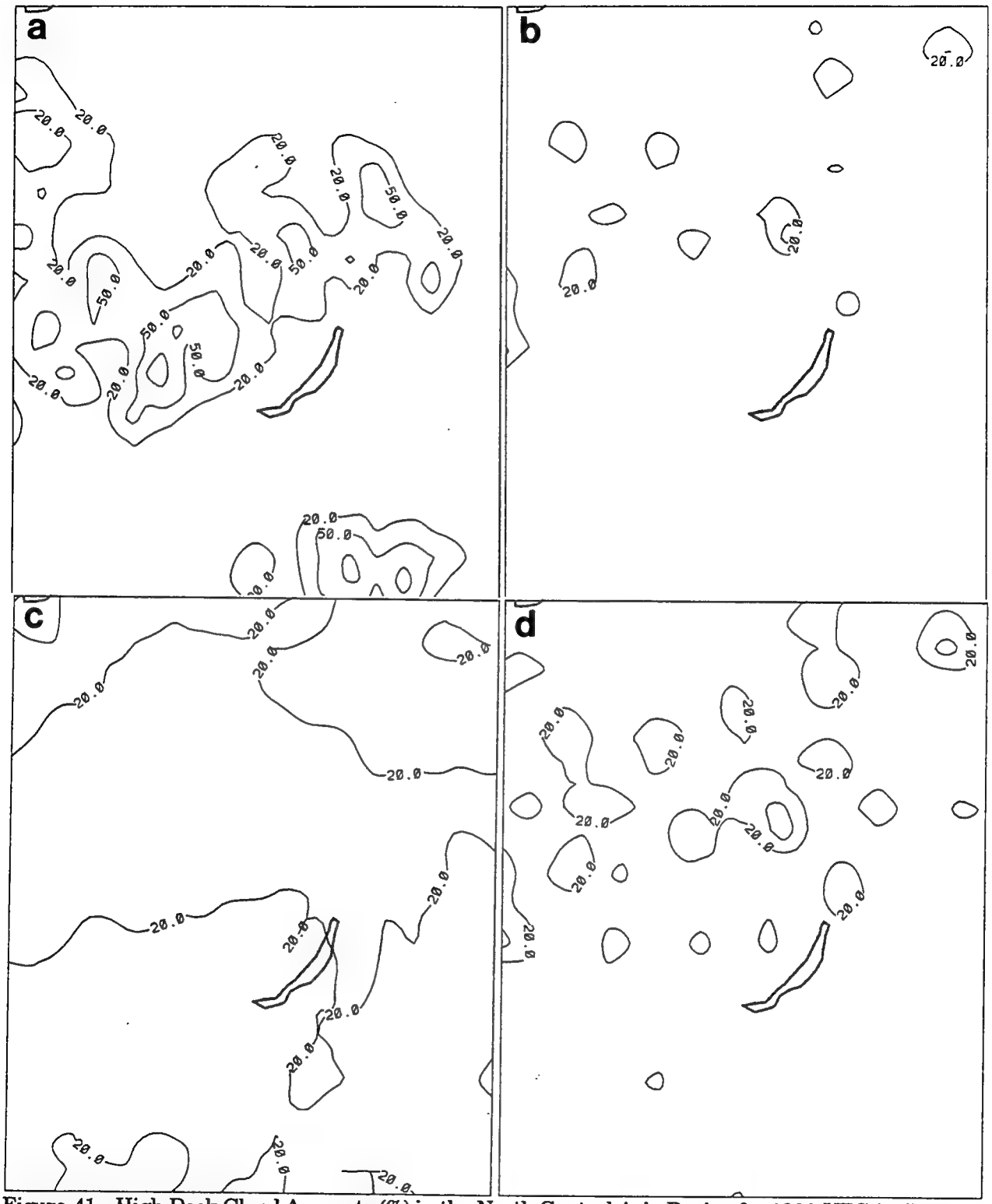

Figure 41. High Deck Cloud Amounts (%) in the North Central Asia Region for 1200 UTC 23 January 1991 from (a) Transformed RTNEPH, (b) 48-Hour CCA Forecast, (c) 48-Hour MLR(RTN) Forecast, (d) 48-Hour MLR(CCA-RTN) Forecast.

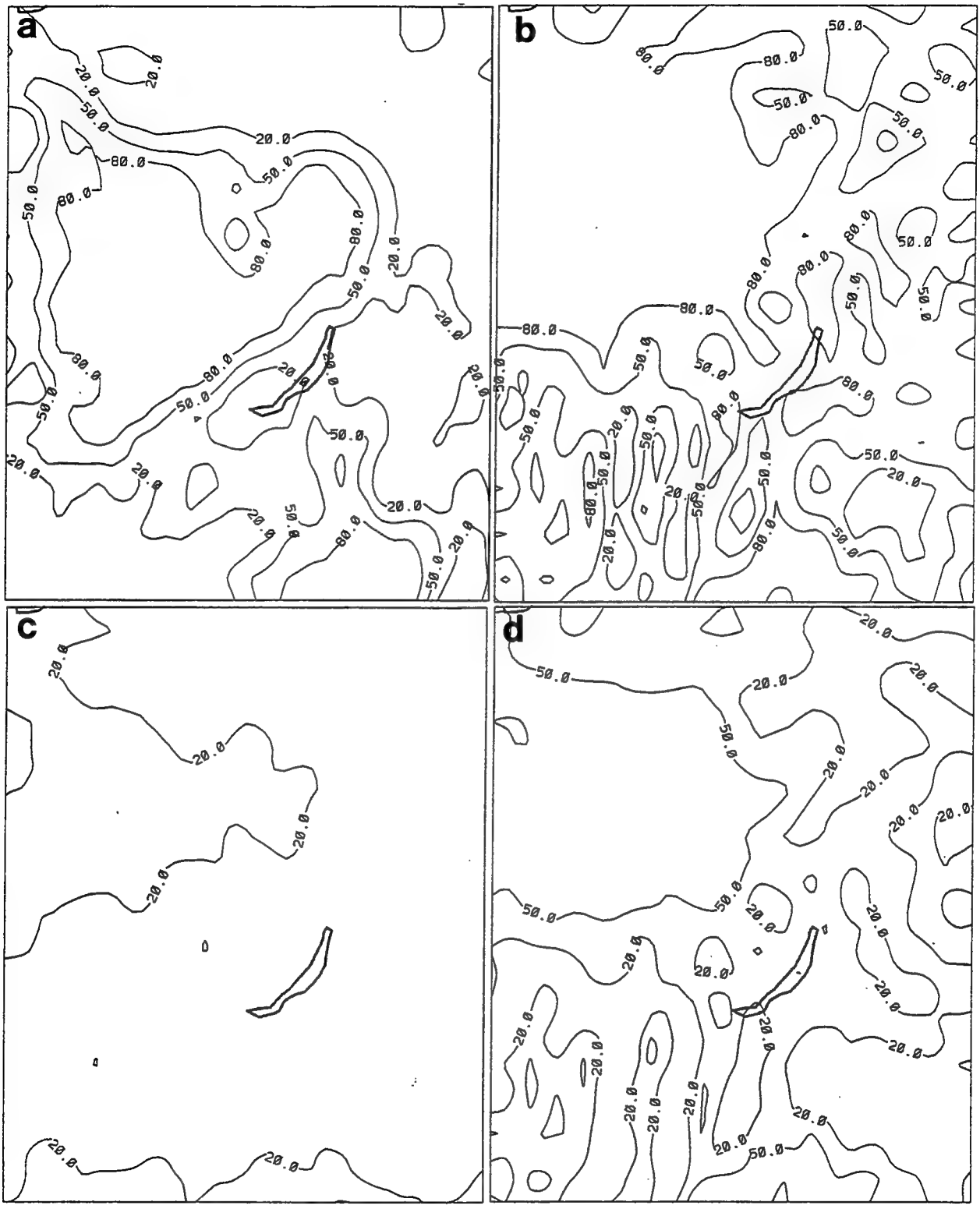

Figure 42. Same as in Figure 41 for Middle Deck Cloud Amounts.

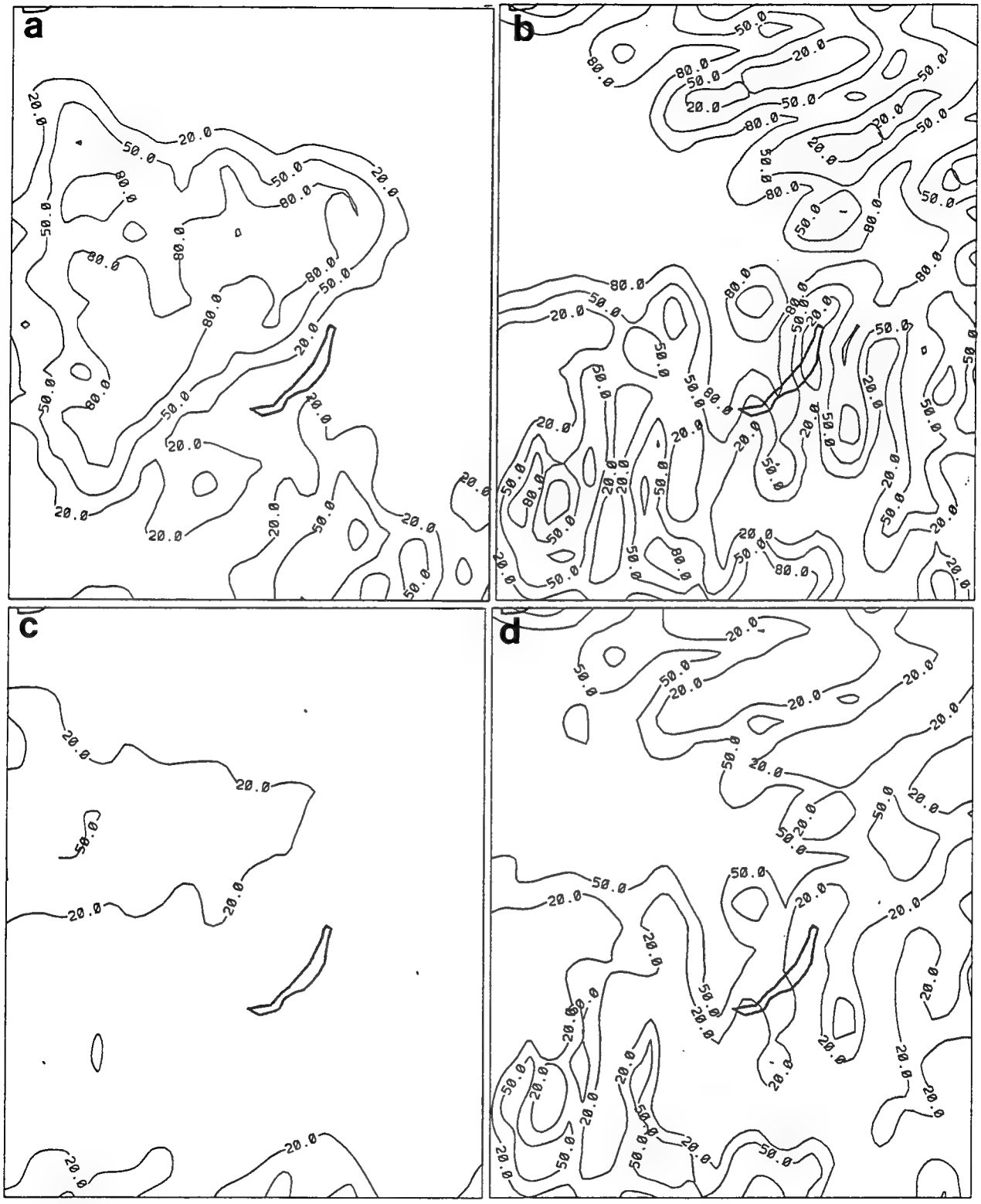

Figure 43. Same as in Figure 41 for Low Deck Cloud Amounts.

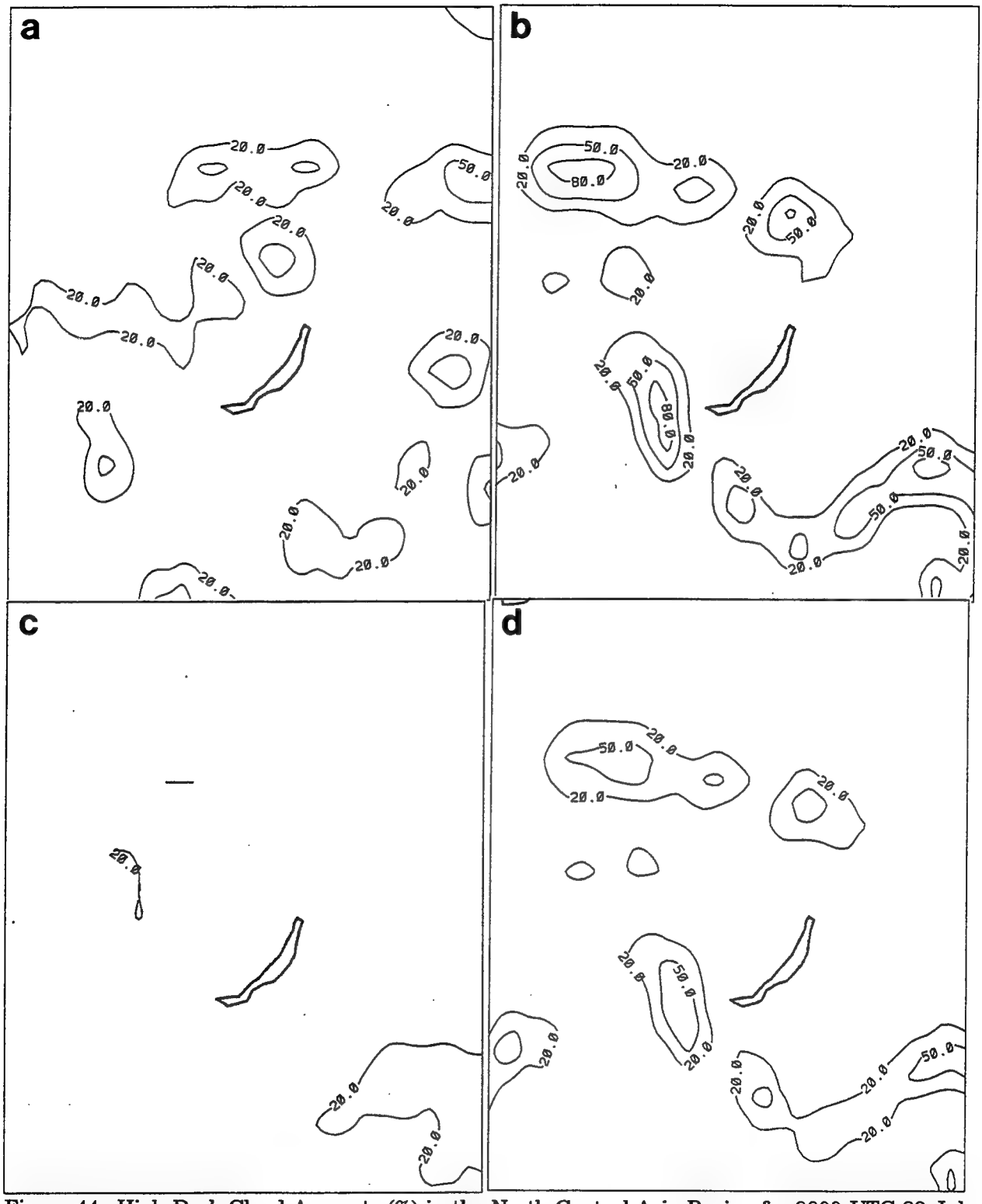

Figure 44. High Deck Cloud Amounts (%) in the North Central Asia Region for 0000 UTC 23 July 1991 from (a) Transformed RTNEPH, (b) 48-Hour CCA Forecast, (c) 48-Hour MLR(RTN) Forecast, (d) 48-Hour MLR(CCA-RTN) Forecast.

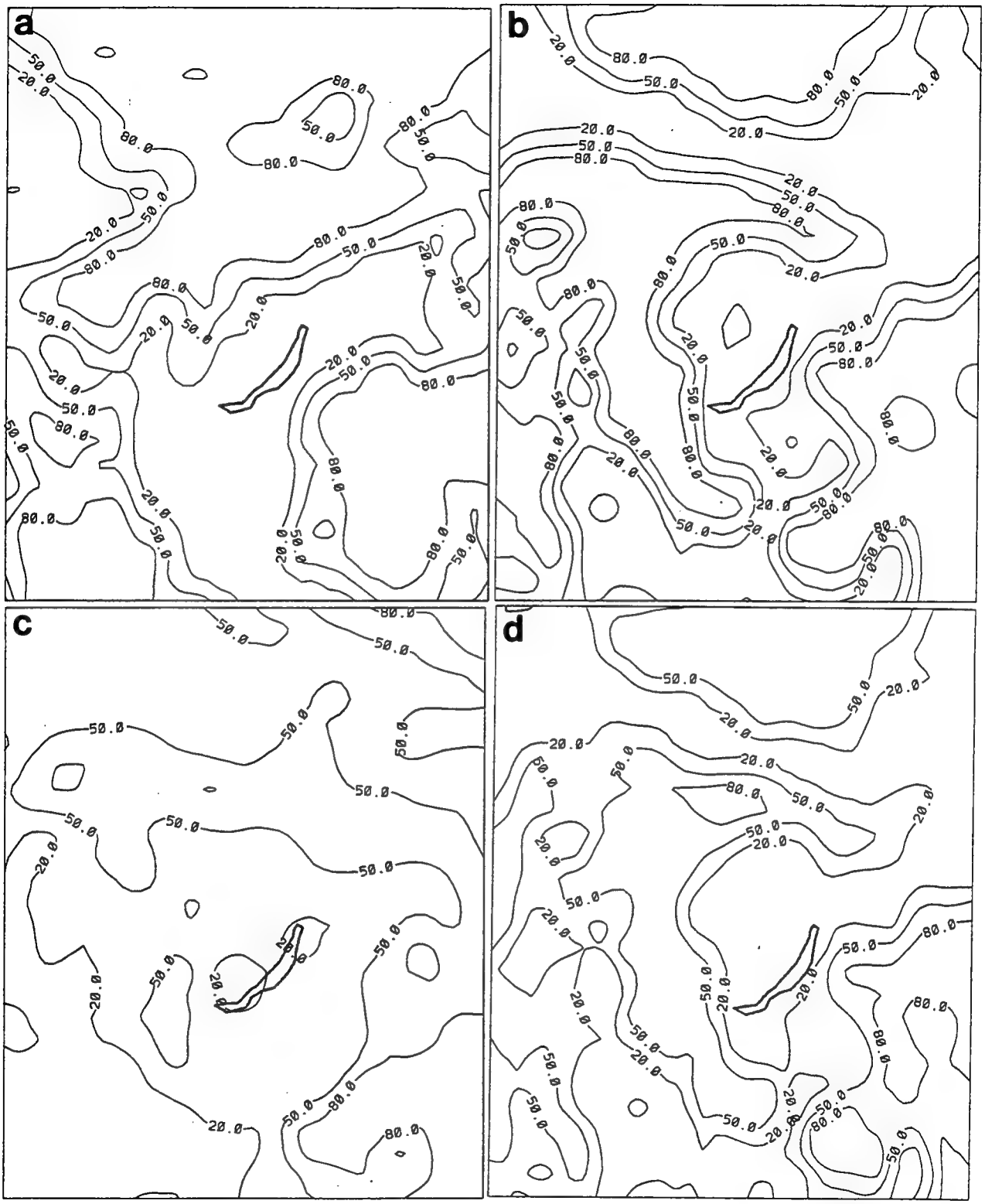

Figure 45. Same as in Figure 44 for Middle Deck Cloud Amounts.

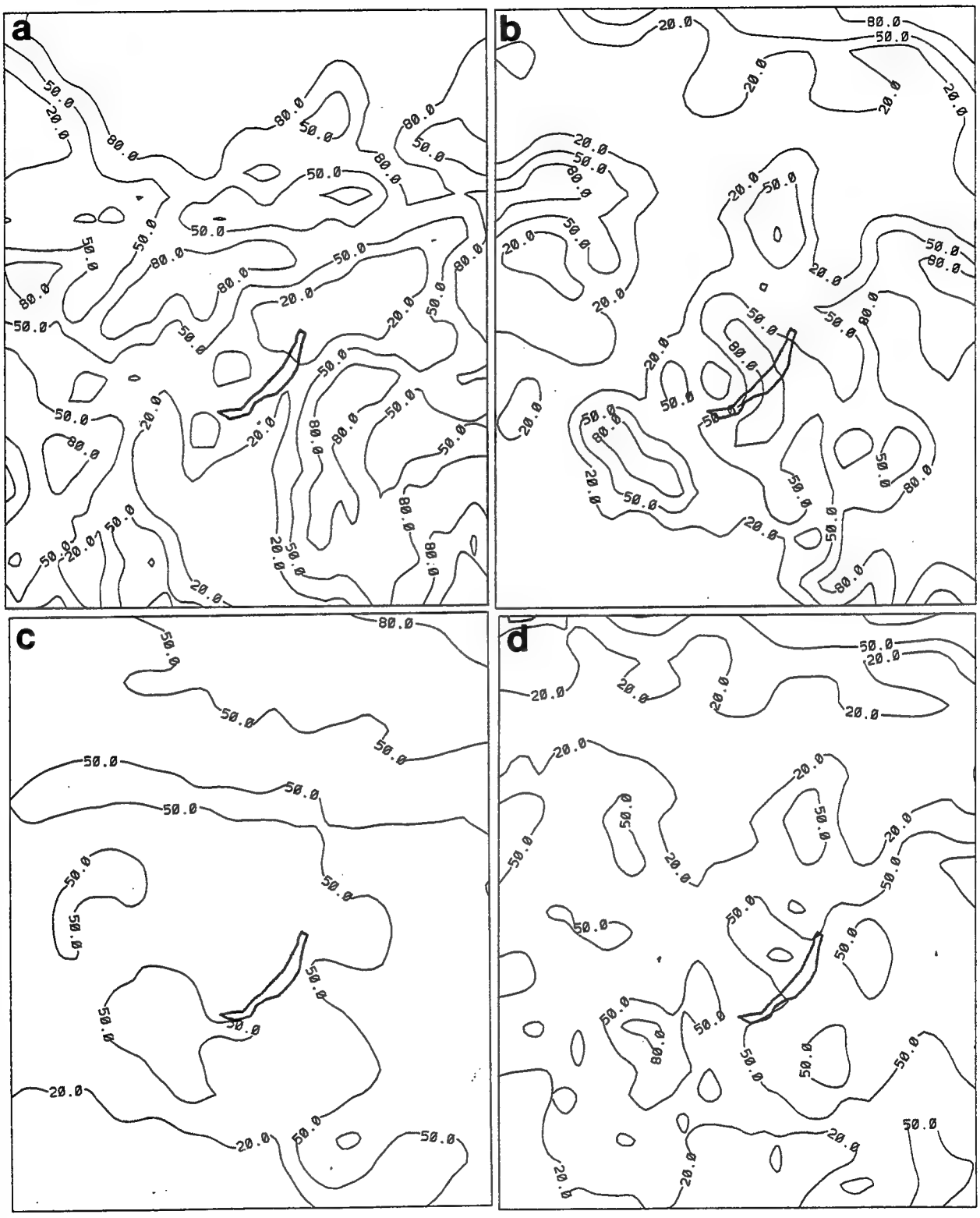

Figure 46. Same as in Figure 44 for Low Deck Cloud Amounts.

The patterns of middle deck cloudiness in the 12-hour (Figure 34) and 48-hour (Figure 45) CCA and MLR(CCA-RTN) are again quite similar to each other at both forecast times. At both times, MLR(CCA-RTN) has less cloudiness by one contour interval (30 percent) than CCA in the CCA's northern and central regions of >80 percent. This suggests an attempt in the MLR(CCA-RTN) to reduce the RMSE by reducing the extremes of cloud amount rather than by relocating them. It appears that the humidity distribution dictates the cloud locations in both schemes, and that the only significant effect that the MLR scheme can have is to change the amounts. Then, as the errors in humidity distribution increase with forecast time, the cloud amount RMSEs grow from the greater mislocation of cloudiness. MLR(RTN) mitigates this growth of RMSE by reducing the extremes of cloudiness to better conform to the predictand. However, MLR(RTN) also is strongly influenced by humidity distribution. This influence wanes in time as is evidenced by a comparison of Figures 34 and 45.

The RTNEPH low cloud distribution in this case is much more irregular than the middle cloud. Yet this greater degree of irregularity is not represented well in any of the forecasts (Figures 35 and 46)--they are all much smoother. The 12-hour CCA and MLR(CCA-RTN) look very similar, but their 48-hour counterparts do not. The CCA depictions and the 12-hour MLR(CCA-RTN) make the areas of extremes of cloudiness (<20 percent and >80 percent) too large in the SW, SE, and northern sections of the region. They do not create a large enough clear area immediately to the south and NE of Lake Baikal. Probably because of the very irregular reference cloud scene and the irregular RH distribution associated with it, the 48-hour MLR(CCA-RTN) reduces most of the extremes of cloudiness to produce a much more bland cloud scene than CCA. MLR(RTN) contrasts with the other two schemes in this case by maintaining a 20-50 percent cloud amount band across the north central portion of the region, where the others have a large <20 percent swath. As in other cases, the extremes are somewhat reduced in the 48-hour MLR(RTN) compared to its 12-hour depiction. This may yield lower RMSEs than the other schemes at 48 hours, but also leads to a graphically less useful forecast.

4.7.3 0000 UTC 23 JULY 1991 SOUTHEAST ASIA CASE

The final example depicts the 48-hour forecast cloud diagnoses for Southeast Asia over land, discussed in Section 4.4.6 for 12-hour forecasts. We discuss in the following paragraphs the comparison between the two forecast times as shown in Figures 38-40 and 47-49. As in Section 4.4.6, we limit our discussion to land areas.

The CCA and MLR(CCA-RTN) high cloud figures (Figures 38 and 47) again show a closer resemblance to each other at both forecast times than they do for each scheme between forecast times. The 12-hour figures of all three schemes show a small patch of high cloud over northern SE Asia that lies between the two patches to the north and south in the RTNEPH. The 48-hour figures show cloudiness along the SE coastline of the SE Asia peninsula that lies between the two cloud masses in the RTNEPH (over central SE Asia and over the Malay Peninsula). Mislocation of high cloudiness occurred in both 12- and 48-hour forecast diagnoses.

RTNEPH cloudiness is dense in the middle deck over all but the NE corner of the region. Interestingly, the 48-hour forecasts of all three schemes (Figure 48) leave more of SE Asia heavily clouded (and thus agree better with RTNEPH) than their 12-hour forecasts. The possibility that this is due to a growing moist bias with forecast time is unlikely (unless it is a very local effect) because of the tendency of the PL-92 model to produce little systematic change in RH on the zonal mean in this latitude domain (see Figure 16). Because of the rather uniform distribution of dense cloud over most of SE Asia in the RTNEPH, the 48-hour MLR(RTN) (usually the most bland of the forecasts) is the most accurate forecast depiction in this case.

As over north central Asia for this date and time, the observed low cloudiness (Figures 40 and 49) shows much greater spatial variance than does middle cloud. This makes it harder to visually judge overall accuracy of forecast depictions. We can say that, in this case, the MLR schemes do not noticeably reduce the cloudiness extremes from 12 hours to 48 hours. This is a departure from the trends we saw in the mid-latitude cases. We cannot conclude on the basis of a single case that tropical behavior of the MLR schemes is different from that in the extratropics.

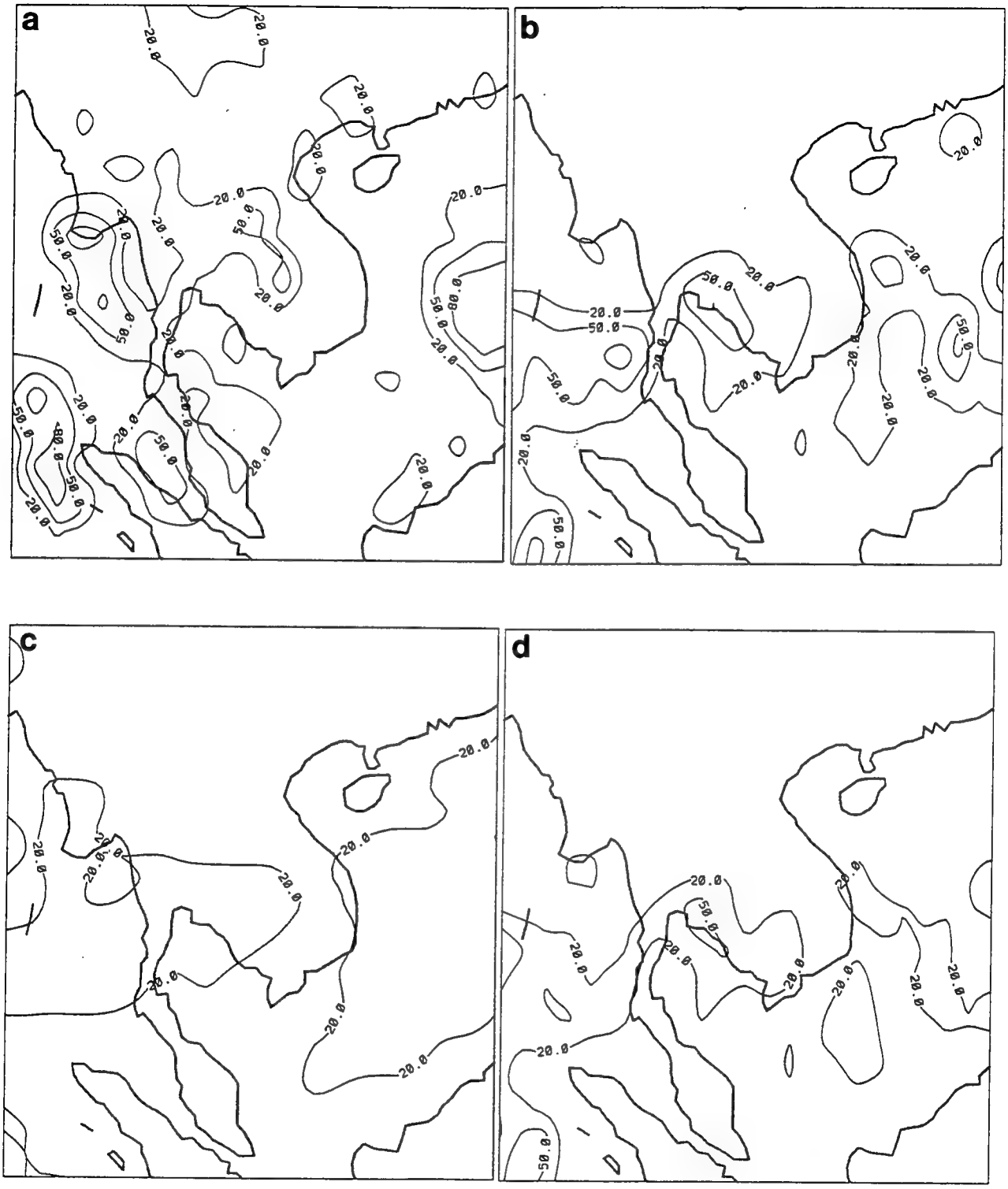

Figure 47. High Deck Cloud Amounts (%) in the Southeast Asia Region for 0000 UTC 23 July 1991 from (a) Transformed RTNEPH, (b) 48-Hour CCA Forecast, (c) 48-Hour MLR(RTN) Forecast, (d) 48-Hour MLR(CCA-RTN) Forecast.

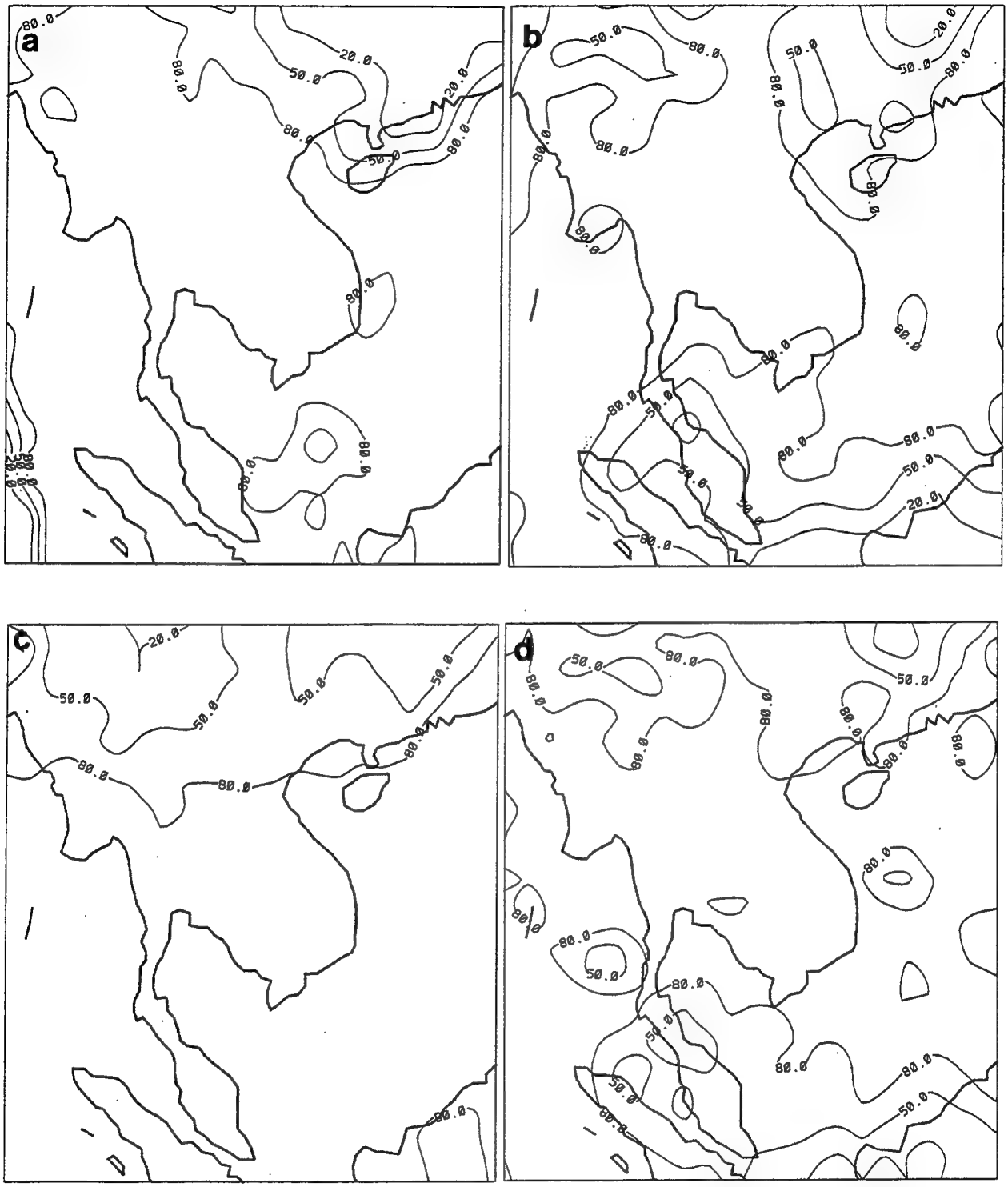

Figure 48. Same as in Figure 47 for Middle Deck Cloud Amounts.

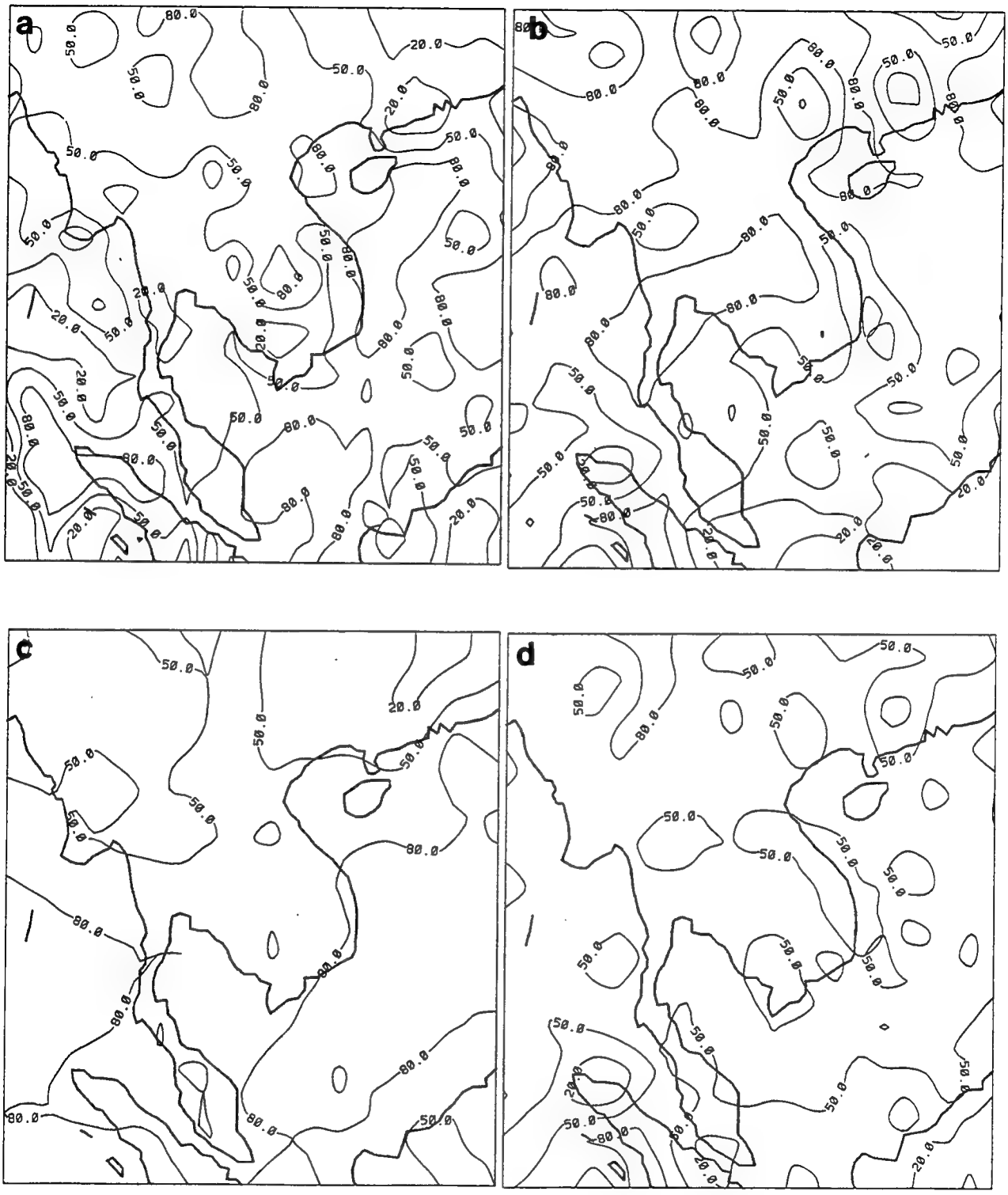

Figure 49. Same as in Figure 47 for Low Deck Cloud Amounts.

5. SUMMARY AND CONCLUSIONS

In this study, we have attempted to investigate the utility of any information derivable from noncloud NWP forecasts to infer layer cloud amount distributions. This effort involved identifying and preparing a suitable source of the predictand (cloud amount), generating and preparing a suitable source of the predictors (NWP variables and geographic information), and combining them to form diagnostic relationships in a statistical approach. The following paragraphs summarize the major findings of this study in each of these components of the problem.

We selected the AFGWC RTNEPH cloud analyses as the source for our cloud amount information. In order to make it spatially compatible with the predictors, we rendered the analyses onto a "transform" grid, characterized by equal-spaced points in the horizontal and three cloud decks (NWP model sigma layer regimes in the vertical). We found a scarcity of independent data available to verify our transform grid cloud amounts objectively. As a result, we used only limited cases of satellite imagery to subjectively confirm the horizontal positioning of cloud features. By comparing global and zonal averages of the transformed RTNEPH with those of earlier investigations of 3DNEPH, we found that we were probably underrepresenting high cloud amounts and overrepresenting middle cloud amounts. We conclude from this that future efforts to use RTNEPH as layer cloud predictand will require an alternative method of assigning middle and high clouds to decks. Also, our choice to ignore high thin clouds in the RTNEPH should be revisited in any future RTNEPH use. Since total cloud amount is likely to be more accurately represented in the RTNEPH than is layer cloud amount, future attempts to diagnose clouds should include total cloud as a predictand.

We found significant differences between transformed RTNEPH cloudiness amounts over land and water surfaces, especially in the winter hemisphere. This suggests that future efforts should determine the utility of regional development and application of cloud diagnosis techniques. The frequency distribution of transformed RTNEPH cloud amount by 5 percent cloud amount categories revealed strong peaks

at 0 percent and 100 percent cloud amount, with a fairly even distribution in between. We conclude that special predictor-predictand relationships must be derived to better fit this nearly bimodal predictand distribution.

In addition to using RTNEPH (RTN) cloud amounts as a predictand, we also used the difference between cloud amounts inferred from the cloud curve algorithm[17] (CCA) method and RTNEPH cloud amounts (CCA-RTN). We derived the CCA RH-to-cloud relationships based on the deck-averaged RH at each transform gridpoint. We found that the use of the deck maximum RH in the CCA relationships produced poorer cloud amount diagnoses.

Deck-average RH was just one of many variables derived from forecasts of the PL GSM NWP model used as potential predictors in our study. Including variable values at 6 hours prior to the valid forecast time, we considered some 99 predictors falling into the categories of dynamic, humidity, geographic, and turbulence. In diagnosing cloud amount from 12-hour forecast variables, we found that deck average RH was consistently the strongest predictor. However, we found that a number of turbulence (such as moist static stability, dry static stability, evaporation rate, boundary layer wind speed) and geographic (such as sine of latitude, hours of darkness before valid time, percent surface water) predictors boosted the total correlation of a linear combination of such predictors over a single most correlated humidity predictor. This suggests that auxiliary (nonhumidity) predictors can play a useful role in diagnosing cloud amount. Interestingly, dynamic variables such as vertical velocity, vorticity, and moisture advection played a rather insignificant role as predictors. Finally, we found that at 48 hours of forecast time, RH and other humidity variables were more poorly correlated with cloud amount. We attribute this lower correlation at longer forecast times to the growth of random (location and magnitude) errors in the PL GSM humidity forecasts. As a result, non-humidity predictors had to play a greater role in the process (many of them nearly maintained their correlation with cloud amount), leading to a total correlation that was only slightly lower at 48 hours than at 12 hours.

We used multiple linear regression (MLR) as the statistical method to relate

linear combinations of the predictor (NWP forecast) variables to RTNEPH cloud amount. It became immediately clear that some modification of the MLR process was necessary to accommodate the nearly bimodal distribution of the predictand. We settled on an increase of the regression slope in the MLR(RTN) application and a withholding of humidity predictors in the MLR(CCA-RTN) as the best compromise between lowest RMSE and highest sharpness in the cloud diagnoses.

We evaluated 12- and 48-hour cloud amount diagnoses from CCA, MLR(RTN), MLR(CCA-RTN) and persistence methods objectively against transformed RTNEPH over a January and July 7-day verification period. None of the diagnostic methods produced significant bias. MLR(RTN) consistently resulted in the lowest mean-squared and mean absolute errors. CCA generally produced the worst scores in both categories, with MLR(CCA-RTN) somewhere in between. Persistence was competitive in skill with MLR(RTN) at 12 hours but not at 48 hours. In 20/20 scores, persistence was competitive with MLR(RTN) at both forecast times. However, in sharpness MLR(RTN) and MLR(CCA-RTN) fell short the essentially perfectly sharp persistence and CCA in the high cloud deck. The MLR schemes produced cloud distributions that generally had only ½ to ⅔ the number of gridpoints with cloud amounts < 20 percent and > 80 percent than RTNEPH had. We concluded from this that the MLR method, even as modified to improve in capturing near-clear and near-overcast cases, sacrificed sharpness to optimize mean-squared error. We conclude that future efforts to statistically relate NWP variables to RTNEPH cloud amounts will require a more deliberate attempt to represent the bimodality of RTNEPH in order to improve sharpness.

This conclusion was confirmed by a subjective evaluation of forecast cloud diagnosis maps in comparison with transform RTNEPH cloud amount maps. Overall, the MLR(RTN) cloud distributions lacked the clear distinctions between nearly clear and nearly overcast present in the RTNEPH. CCA cloud distributions did retain the clarity, but suffered from the mislocation of the cloud features due to humidity forecast errors. These humidity mislocations also negatively influenced the MLR(CCA-RTN) and MLR(RTN) diagnoses. In fact, most CCA and MLR(CCA-RTN)

depictions looked very similar. However, the MLR tended to reduce the extremes of cloudiness in the depictions in an attempt to reduce the RMSE caused by the humidity mislocation. The forecasted humidity distribution dictates the cloud locations in all three diagnostic methods. MLR(RTN) [and to a lesser degree, MLR(CCA-RTN)] tend to mitigate RMSE growth with forecast time (resulting from growth in humidity forecast error) by reducing the cloud amount extremes. This results in a flatter, less well-defined, less useful cloud amount depiction. From these findings, we conclude that, to improve the usefulness of diagnosed cloud distributions, (1) we must improve the NWP model's ability to forecast RH distribution, and (2) we must improve the sharpness of the cloud distribution diagnosed from RH and auxiliary variables.

Future efforts at Phillips Laboratory will concentrate on improving PL GSM humidity forecasts, and improving RMSE and sharpness of cloud distributions diagnosed from them and auxiliary variables. Our ultimate goal in future research is to determine and demonstrate the greatest skill level possible in large-scale NWP model-based cloud diagnosis.

References

1. Callahan, K.P., Kiess, R.B., Lanicci, J.M., and Neu, T.J. (1993) Cloud Analysis and Forecasting at Air Force Global Weather Central Under the Cloud Depiction and Forecasting System II, *Proceedings, Cloud Impacts on DOD Operations and Systems 1993 Conference,* 16-19 November 1993, U.S. Army Topographic Engineering Center, Fort Belvoir, VA, 416-421.

2. Kiess, R.B. and Cox, W. (1988) *The AFGWC Automated Real-Time Cloud Analysis Model,* AFGWC/TN-88-001, Air Force Global Weather Central, Offutt AFB, NE, 82 pp.

3. Hamill, T.M., d'Entremont, R.P. Bunting, J.T. (1992) A Description of the Air Force Real-Time Nephanalysis Model, *Mon. Wea. Rev.,* **7**:288-306.

4. Crum, T.D. (1987) *AFGWC Cloud Forecast Models*, AFGWC/TN-87-001, Offutt AFB, NE, 66 pp.

5. Stobie, J.G. 1986: *AFGWC's Advanced Weather Analysis and Prediction System (AWAPS)*, AFGWC/TN-86-001, Air Force Global Weather Central, Offutt AFB, NE, 98 pp.

6. Isaacs, R.G., Gustafson, G.B., Snow, J.W., and d'Entremont, R.P. (1993) Satellite Cloud Analysis Programs at the Air Force Phillips Laboratory: An Overview -Part 2 Support of Environmental Requirements for Cloud Analysis and Archive (SERCAA), *Proceedings, Cloud Impacts on DOD Operations and Systems 1993 Conference*, 16-19 November 1993, U.S. Army Topography Engineering Center, Ft. Belvoir, VA, 422-427.

7. Geleyn, J.-F. (1981) Some Diagnostics of the Cloud/Radiation Interaction in ECMWF Forecasting Model, *Workshop on Radiation and Cloud-Radiation Interaction in Numerical Modeling*, ECMWF, 15-17 October 1980, 135-162.

8. Slingo, J.M. (1987) The Development and Verification of a Cloud Prediction Scheme for the ECMWF Model, *Quart. J. Roy. Meteor. Soc.,* **113**:899-927.

9. Campana, K., Mitchell, K., Yang, S., Hou, Y., and Stowe, L. (1991) Valiation and Improvement of Cloud Parameterization at the National Meteorological Center, *Preprints, 9th Conference on Numerical Prediction,* 14-19 October 1991, Denver, CO, pp. 131-134.

10. Kiehl, J.T., 1991: Modeling and Validation of Clouds and Radiation in the NCAR Community Climate Model, *ECMWF/WCRP Workshop: Clouds, Radiation Transfer and the Hydrological Cycle,* 249-272.

11. Schattel, J. (1992) *Refinement and Testing of the Radiative Transfer Parameterization in the PL Global Spectral Model,* PL-TR-92-2169, Phillips Laboratory, Hanscom AFB, MA, 27 pp. ADA256840.

12. Mitchell, K. and Hahn, D. (1989) *Development of a Cloud Forecast Scheme for the GL Baseline Global Spectral Model,* GL-TR-89-0343, Geophysics Laboratory Hanscom AFB, MA 152 pp. ADA231595.

13. Rikus, L. and Hart, T. (1988) The Development and Refinement of a Diagnostic Cloud Parameterization Scheme for the BMRC Global Model. *Proceedings, International Radiation Symposium (IRS88),* Lille, France, 14-24 August 1988.

14. Kvamsto, N. (1991) An Investigation of Diagnostic Relations Between Stratiform Fractional Cloud Cover and Other Meteorological Parameters in Numerical Weather Prediction Models. *J. App. Meteor.,* **30:**200-216.

15. Sheu, R.-S., and Curry, J.A. (1992) Interactions Between North Atlantic Clouds and the Large-Scale Environment, *Mon. Wea. Rev.,* **120:**261-278.

16. Smith, R.N.B. (1990) A Scheme for Predicting Layer Clouds and Their Water Content in a General Circulation Model, *Quart. J. Roy. Meteor. Soc.,* **116:**435-460.

17. Trapnell, R.N., Jr. (1992) *Cloud Curve Algorithm Test Progrram,* PL-TR-92-2052, Phillips Laboratory, Hanscom AFB, MA. ADA253918.

18. Zivkovic, M. and Louis, J.-F. (1991) A Cloud Parameterization Based on Cluster Analysis, Preprints, *Ninth Conference on Numerical Weather Prediction,* 14-19 October 1991, Denver, Amer. Meteor. Soc., 132-134.

19. Bao, J.-W. (1990) *Accuracy of Cloud Forecasts Obtained from Mesoscale Numerical Weather Prewdiction Model,* M.S. Thesis, Pennsylvania State University, August 1990, 128 pp.

20. Krishnamurti, T.N., Bedi, H.S., Ingles, K., Weiner, A., Kuma, K., Campana, K.A., and Kimoto, M. (1988) Comparison of Cloud Cover from Global Models and ISCCP Data Sets, Report prepared for the Fourth Session of the CAS/JSC-WGNE, Toronto, 16-30 September 1988, 67 pp.

21. Wu, X. (1990) *Assimilation of ERBE Data with Nonlinear Programming Technique to Improve Cloud Cover Diagnostics,* Ph.D. Thesis, University of Wisconsin, Madison, University Microfilms #DA9034859, 113 pp.

22. Cianciolo, M.E. (1993) *Short-Range Cloud Amount Forecasting With Model Output Statistics,* PL-TR-93-2205, Phillips Laboratory, Hanscom AFB, MA, 154 pp. ADA274769.

23. Glahn, H.R. (1985) Statistical Weather Forecasting, In *Probability, Statistics, and Decision Making in the Atmospheric Sciences,* A.H. Murphy and R.W. Katz, eds., Westview Press, Boulder, CO, 289-335.

24. Glahn, H.R. and Lowry, D.A. (1972) The Use of Model Output Statistics (MOS) in Objective Weather Forecasting, *J. Appl. Meteor.,* **11:**1203-1211.

25. Murphy, A.H. (1985) Probabilistic Weather Forecasting, In *Probability, Statistics, and Decision Making in the Atmospheric Sciences,* A.H. Murphy and R.W. Katz, eds., Westview Press, Boulder, CO, 337-377.

26. Reinke, D.L., Eis, K.E., Forsythe, J.M., Combs, C.L., and Vonder Haar, T.H. (1993) Comparison of the Real Time Nephanalysis (RTNEPH) with High Resolution Satellite Cloud Climatology (HRSCC), *Proceedings, Cloud Impacts on DOD Operations and Systems 1993 Conference,* 16-19 November 1993, U.S. Army Topographic Engineering Center, Ft. Belvoir, VA, 440-445.

27. Smagorinsky, J. (1960) On the Dynamical Prediction of Large-Scale Condensation by Numerical Methods, *Geophysical Monograph No. 5,* 71-78.

28. Hortal, M. (1990) Use of Reduced Gaussian Grids in Spectral Models, *ECMWF Newsletter,* Number 52, December 1990, European Centre for Medium-Range Weather Forecasts, 3-7.

29. Henderson-Sellers, A. (1986) Layer Cloud Amounts for January and July 1979 from 3D-Nephanalysis, *J. Clim. Appl. Meteor.,* **25:**118-132.

30. Norquist, D.C., Yang, C., Chang, S., and Hahn, D.C. (1992) *Phillips Laboratory Global Spectral Numerical Weather Prediction Model,* PL-TR-92-2225, Phillips Laboratory, Hanscom AFB, MA, 154 pp. ADA267239.

31. Nehrkorn, T., Hoffman, R.N. Louis, J.-F., and Zivkovic, M. (1992) *An Enhanced Global Spectral Model,* PL-TR-92-2011, Phillips Laboratory, Hanscom AFB, MA. ADA25142.

32. Derickson, R.G. and Cotton, W.R. (1977) *On the Use of Finite Taylor's Series Approximations to Certain Exponential and Power Functions Employed in Cloud Models,* Atmospheric Science Paper 268, Colorado State University, Fort Collins, CO.

33. Murray, F.W. (1967) On the Computation of Saturation Vapor Pressure, *J. Appl. Meteor.,* **6:**203-204.

34. Norquist, D.C. and Chang, S.S. (1994) Diagnosis and Correction of Systematic Humidity Error in a Global Numerical Weather Prediction Model, accepted for publication in *Mon. Wea. Rev.*